

volume 43 number 4

december 2021

the sydney law review

articles

- Out of Limbo and Into the Light: A Case for Status Resolution
for Undocumented Migrant Workers on Farms
– *Joanna Howe* 433

- Keeping the (Good) Faith: Implications of Emerging Technologies
for Consumer Insurance Contracts
– *Zofia Bednarz and Kayleen Manwaring* 455

- Fighting the System: New Approaches to Addressing Systematic
Corporate Misconduct
– *Samuel Walpole and Matt Corrigan* 489

- The “Constitutional” Value of the *Racial Discrimination Act 1975* (Cth)
– *Alice Taylor* 519

- Accuracy, Utility and Gateways: Justifications(?) for Controlling the
Use of Surrounding Circumstances in Contractual Interpretation
– *NA Tiverios* 547

before the high court

- Liability for Workplace Psychiatric Injury and Vicarious Trauma:
Kozarov v Victoria
– *Kylie Burns* 575

case note

- Love v Commonwealth*: The Section 51(xix) Aliens Power and a
Constitutional Concept of Community Membership
– *Mischa Davenport* 589

EDITORIAL BOARD

Jacqueline Mowbray (Editor)

Kimberlee Weatherall (Editor)

Celeste Black

Ben Chen

David Hamer

Emily Hammond

Jason Harris

Grant Hooper

Ghena Krayem

Kristin Macintosh

Michael Sevel

Belinda Smith

Yane Svetiev

STUDENT EDITORIAL COMMITTEE

Nina Dillon Britton

Rishabh Khanna

Rebekah Ord

Harriet Holowell

Rachael Li

Grace Qiu

Wendy Hu

Tom Manousaridis

Anuki Suraweera

Alan Zheng

Before the High Court Editor: Emily Hammond

Book Review Editor: Michael Sevel

Publishing Manager: Cate Stewart

Correspondence should be addressed to:

Sydney Law Review

Sydney Law School

Building F10, Eastern Avenue

UNIVERSITY OF SYDNEY NSW 2006

AUSTRALIA

Email: sydneylawreview@sydney.edu.au

Website and submissions: [<https://sydney.edu.au/law/our-research/publications/sydney-law-review.html>](https://sydney.edu.au/law/our-research/publications/sydney-law-review.html)

For hardcopy subscriptions outside North America:

sydneylawreview@sydney.edu.au

For hardcopy subscriptions in North America: info@gaunt.com

The *Sydney Law Review* is a refereed journal.

© 2021 Sydney Law Review and authors.

ISSN 0082-0512 (PRINT)

ISSN 1444-9528 (ONLINE)

Out of Limbo and into the Light: A Case for Status Resolution for Undocumented Migrant Workers on Farms

Joanna Howe*

Abstract

The Australian horticulture industry has endemic labour challenges, both in terms of poorly managed and distributed labour supply and a systemic problem of non-compliance with labour standards. A core component of both problems is the entrenched reliance on undocumented migrant workers. This article examines the extent of this reliance and considers policy solutions to address it. In particular, it proposes a model for the one-off status regularisation of undocumented migrant workers in the Australian horticulture industry.

Please cite this article as:


Joanna Howe, 'Out of Limbo and into the Light: A Case for Status Resolution for Undocumented Migrant Workers on Farms' (2021) 43(4) *Sydney Law Review* 433.



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC 4.0).

As an open access journal, unmodified content is free to use with proper attribution. Please email sydneylawreview@sydney.edu.au for permission and/or queries.

© 2021 Sydney Law Review and author. ISSN: 1444-9528

* Associate Professor in Law, Adelaide Law School, University of Adelaide, South Australia, Australia; 2020–21 Member, Ministerial Advisory Council on Skilled Migration Email: joanna.howe@adelaide.edu.au; ORCID iD:  <https://orcid.org/0000-0003-0195-9734>. The author is indebted to Irene Nikoloudakis for her invaluable research assistance.

I Introduction

In a watershed moment for the Australian horticulture sector, the *National Agriculture Workforce Strategy* strongly recommended that the Australian Government introduce a one-off regularisation process for undocumented farm workers.¹ A month later, the National Party Federal Conference voted unanimously in favour of status resolution.² The presence of undocumented workers, which the *National Agriculture Workforce Strategy* estimates is between 60,000 and 100,000 workers,³ is the dark underbelly of a sector reliant on an overseas workforce to pick fresh fruit and vegetables. Undocumented workers form a critical part of the harvest workforce. Given complex supply chains transiting fresh fruit and vegetables from the farm to the consumer, undocumented workers are invisible to the Australian public. Yet, it is likely that all Australians, at one time or another, have purchased fruit and vegetables harvested through the labour of undocumented workers.⁴

This article contends that there are strong arguments in favour of status resolution for undocumented migrants on farms. First, this reform will help the sector's labour challenges. Although the horticulture sector has struggled for many years with difficulties in recruiting and retaining harvest workers, this was exacerbated when border restrictions were introduced due to the COVID-19 pandemic.⁵ Although regularising the status of undocumented workers will not increase the farm labour pool in Australia, it will improve the mobility of these workers. It will enable undocumented migrants to cross state borders and enable ethical growers, who previously did not have access to undocumented workers, to employ this substantial group of workers. Status regularisation is not a silver bullet for labour shortages on farms, but will make a sizable contribution to addressing them.

The second reason for a one-off regularisation process for undocumented farm workers is that it is needed to remove the susceptibility of this group to

¹ J Azarias, R Nettle and J Williams (National Agricultural Labour Advisory Committee), *National Agricultural Workforce Strategy: Learning to Excel* (Department of Agriculture, Water and the Environment (Cth), December 2020) 190 ('*National Agricultural Workforce Strategy*'). In this article, undocumented migrant farm workers refers to migrants working in the industry without an entitlement to work. These include: migrants on visas without work rights (such as tourists); migrants whose visas have expired; and migrants with a valid visa with work rights, but who work in breach of a condition of their visa.

² Anne Webster, 'National Party Federal Conference Supports Status Resolution' (Media Release, 30 March 2021) <<https://www.annewebster.com.au/media/national-party-federal-conference-supports-status-resolution/>>.

³ *National Agricultural Workforce Strategy* (n 1) 206, citing J Howe and A Singh, *COVID-19 and Undocumented Workers in the Australian Horticulture Industry* (University of Adelaide Law Research Paper No. 2020-137, 2020).

⁴ Jack Snape, 'Australian Berry and Citrus Farms Most at Risk of Having Slaving-Like Working Conditions, Woolworths Modern Slavery Report Finds', *ABC News* (online, 26 November 2020) <<https://www.abc.net.au/news/2020-11-26/berries-grapes-modern-slavery-coles-woolworths/12918122>>; Sarina Locke, 'Debt Bondage for Workers in Australian Horticulture Akin to Slavery, Inquiry Hears', *ABC News* (online, 19 October 2017) <<https://www.abc.net.au/news/rural/2017-10-19/debt-bondage-in-horticulture-sector-akin-to-slavery-in-australia/9057108>>.

⁵ Kath Sullivan, 'Farmers Fear Worker Shortage due to COVID-19 Restrictions despite Rising Unemployment', *ABC News* (online, 30 July 2020) <<https://www.abc.net.au/news/rural/2020-07-30/farm-labour-shortage-feared-due-to-coronavirus-controls/12504802>>.

exploitation. There have been myriad media exposés, inquiries and reports exposing wage theft and other forms of exploitation on farms.⁶ Poor labour standards on farms is made possible because undocumented workers are vulnerable to exploitation as they have no right to work in Australia and typically access farm work through unscrupulous contractors. If undocumented migrant workers are incentivised to regularise their status, they will be less likely to tolerate exploitation, more likely to join unions, and more likely to report exploitation to the Fair Work Ombudsman, Australia's workplace regulator, and other support services.

The third reason to introduce a one-off regularisation procedure is because there is no legitimate alternative that can address the labour crisis on Australian farms. The alternative to one-off regularisation is a policy of detection and deportation of undocumented workers. There has been a resounding public policy failure from successive federal governments over two decades to address the substantial presence of undocumented workers on farms. Despite the deployment of considerable resources and the development of a special taskforce in 2015, efforts to detect and deport undocumented workers have been so ineffective that undocumented workers have grown to become a key and significant part of the sector's workforce. Additionally, although introducing national labour hire licensing⁷ will have some bearing on improving labour standards in the sector, it will not address the substantial presence of undocumented workers.

Part II of this article examines the farm labour crisis and considers the arguments in favour of status resolution in terms of both meeting labour needs and addressing endemic exploitation. Part III examines the profile and prevalence of undocumented migrants in the horticulture sector and reviews government efforts to detect and deport undocumented farm workers. Part IV develops a skeletal framework for a one-off status resolution process and considers the arguments against status resolution and whether other reforms can address the labour crisis without the need for status resolution.

⁶ See, eg, 'Slaving Away: The Dirty Secrets Behind Australia's Fresh Food', *Four Corners* (Australian Broadcasting Corporation, 2015) <<https://www.abc.net.au/4corners/slaving-away-promo/6437876>>; Tom Stayner, 'New Report Warns Some Migrant Workers Paid Less than \$2 an Hour on Australian Farms', *SBS News* (online, 19 March 2021) <<https://www.sbs.com.au/news/new-report-warns-some-migrant-workers-paid-less-than-2-an-hour-on-australian-farms>>; 'More than \$1m in Lost Wages Recovered for 2,500 Fruit Farm Workers', *The Guardian* (online, 22 November 2018) <<https://www.theguardian.com/australia-news/2018/nov/22/more-than-1m-in-lost-wages-recovered-for-2500-fruit-farm-workers>>; Nick McKenzie and Richard Baker, 'Fruits of their Labour', *The Sydney Morning Herald* (online, 2016) <<https://www.smh.com.au/interactive/2016/fruit-picking-investigation/>>; Fair Work Ombudsman, *Harvest Trail Inquiry: A Report on Workplace Arrangements along the Harvest Trail* (Report, 2018); Joanna Howe, Stephen Clibborn, Alexander Reilly, Diane van den Broek and Chris F Wright, *Towards a Durable Future: Tackling Labour Challenges in the Australian Horticulture Industry* (Report, January 2019) <<https://www.sydney.edu.au/content/dam/corporate/documents/business-school/research/work-and-organisational-studies/towards-a-durable-future-report.pdf>> ('*Durable Future Report*'); Senate Education and Employment References Committee, Parliament of Australia, *A National Disgrace: The Exploitation of Temporary Work Visa Holders* (Report, March 2016).

⁷ Australian Government, *Australian Government Response: Report of the Migrant Workers' Taskforce* (March 2019) 3 <https://www.ag.gov.au/sites/default/files/2020-03/government_response_to_the_migrant_workers_taskforce_report.pdf>.

II The Farm Labour Crisis

The Australian horticulture sector is facing an urgent and immediate labour crisis. There are two dimensions to this crisis. The first concerns an inability to access sufficient labour. The geographically dispersed and seasonal nature of the horticulture sector means that these labour challenges are not universal across the sector and manifest differently during the harvest for certain commodities and at different times in the year. The second dimension of the labour crisis concerns endemic exploitation. This Part examines these ongoing labour challenges, which have a long history of affecting horticultural labour supply, but have been exacerbated by the COVID-19 pandemic. It is important to understand these dimensions before examining the prevalence and profile of undocumented migrant workers on farms.

Farm Labour Supply Challenges

With international borders closed as a result of the COVID-19 pandemic, a significant decline in the supply of overseas labour has caused labour shortages. There have been multiple reports that growers have been ploughing fruit and vegetables back into the field or not harvesting produce because of an inability to find workers.⁸ As at February 2021, the National Lost Crop Register, an initiative designed by Growcom for growers to report anonymously the costs of ongoing labour shortages (which are beyond financial), estimated losses to have surpassed \$45 million.⁹ Further, economic modelling predicted that the industry would be short of up to 26,000 casual workers between November 2020 and June 2021.¹⁰ Notably, as this modelling was based on the assumption that the reopening date for international borders would be March 2021, it may have underestimated the shortfall given that borders remain at least partially closed for the remainder of 2021 as at the time of writing.

⁸ See, eg, Jessica Hayes, 'Farm Worker Shortage Could Mean More Expensive, Lower Quality Fruit and Vegetables This Summer', *ABC News* (online, 22 November 2020) <<https://www.abc.net.au/news/2020-11-22/christmas-lunch-to-cost-australians-more-in-2020/12901704>>; Natalie Kotsios, 'Picking Up the Pieces', *The Weekly Times* (Melbourne, 25 November 2020); Jon Daly and Conor Byrne, 'COVID-19 is Pushing Up Prices for NT Mangoes, but Growers are Battling to Pick Them', *ABC News* (online, 25 October 2020) <<https://www.abc.net.au/news/rural/2020-10-25/covid-19-is-driving-labour-shortages-and-higher-prices-nt-mango/12806046>>; Ellen Ransley, 'Overseas Labour Shortage Causes Fruit, Vegie Production to Fall, Prices to Rise', *News.com.au* (online, 3 March 2021) <<https://www.news.com.au/finance/money/costs/overseas-labour-shortage-causes-fruit-vegie-production-to-fall-prices-to-rise/news-story/2a3ae82dc34a583725fc6e4730a74583>>.

⁹ James Liveris, 'National Lost Crop Register Surpasses \$45 million in Losses at Farmgate Value Due to Worker Shortage', *ABC Rural* (online, 9 February 2021) <<https://www.abc.net.au/news/rural/2021-02-09/national-lost-crop-register-surpasses-45-million-in-losses/13132274>>. See also Growcom, 'Fruit and Veg Losses Top \$38 million' (Media Release, 8 January 2021) <<https://www.growcom.com.au/2021/01/08/fruit-and-veg-losses-top-38-million/>>.

¹⁰ Ernst & Young, *Seasonal Horticulture Labour Demand and Workforce Study: Public Report* (September 2020) 14.

1 *A Snapshot of the Farm Labour Market before COVID-19*

The number of workers in the horticulture sector is reported by the Australian Bureau of Agricultural and Resource Economics and Sciences ('ABARES') to be approximately 135,100.¹¹ Adequate data are not collected on the volume of workers required in each occupation within this workforce.¹² Nevertheless, it would appear that pickers, packers and graders comprise the largest group of workers in the horticulture sector. ABARES states that '[l]abourers accounted for around three-quarters of the workforce on vegetable farms and fruit and nut farms'.¹³ This means there are approximately 101,000 workers in the industry.¹⁴

Prior to the COVID-19 pandemic, temporary migrants were the main source of farm labour for harvesting fruit and vegetables, accounting for three-quarters of the workforce.¹⁵ In the year before the pandemic, the number of Working Holiday Makers ('WHMs') in the horticulture sector was steadily growing, with 43,219 second year visa grants for WHMs employed in 'specified work' for an 88-day period during the first year of their visa.¹⁶ Approximately 80% of these second year visa grants (over 36,000 WHMs) earned their visa extension through employment on farms.¹⁷ Although it is likely that some WHMs perform horticultural work outside of the 88 days, this is likely to be a small group. Notably, ABARES states that '[i]n ineligible postcode regions, backpackers made up just 9% of the workforce of vegetable farms'.¹⁸ In 2018–19, the Seasonal Worker Programme ('SWP') grew significantly, with approvals for 12,202 workers from Pacific countries to enter Australia for seasonal harvest work.¹⁹

¹¹ According to data provided by the Australian Government Department of Agriculture, Water and the Environment there were 135,100 workers on fruit, grape, nut and vegetable farms in 2020–21, down from 146,200 in 2019–20: see Department of Agriculture, Water and the Environment (Cth), Australian Bureau of Agricultural and Resource Economics and Sciences ('ABARES'), *Labour Use in Australian Agriculture: Analysis of Survey Results* (Web Page, 25 November 2021) <<https://www.awe.gov.au/abares/research-topics/labour>>.

¹² *National Agricultural Workforce Strategy* (n 1) ch 8 (workforce data and information).

¹³ Niki Duffy, Peter Martin and Shiji Zhao, *Demand for Farm Workers: ABARES Farm Survey Results 2018* (ABARES Research Report 19.10, September 2019) 13.

¹⁴ There are known difficulties in data collection in this area. On the one hand, data published by the Australian Bureau of Statistics ('ABS') is likely to undercount substantially the total horticultural workforce as it does not include overseas workers and contract workers, both of whom are key components of the horticultural labour force. ABS data are also unlikely to include undocumented workers. However, on the other hand, ABS data do potentially lead to workers being counted multiple times as seasonal production on some farms means workers can work on multiple farms. ABARES acknowledges the limitations of the available data and states that its figures are based on its 'best estimates' of labour use on farms drawing on ABS data on farm numbers and ABARES farm survey data: see Peter Martin, Lucy Randall and Tom Jackson, *Labour Use in Australian Agriculture* (ABARES Research Report 20.20, December 2020) 5 ('Key Caveats and Assumptions') <<https://doi.org/10.25814/gjyp-7g19>>.

¹⁵ Haydn Valle, Niki Millist and David Galeano, *Labour Force Survey* (ABARES report to the Department of Agriculture and Water Resources (Cth), May 2017).

¹⁶ Department of Home Affairs (Cth), Submission No 82 to Select Committee on Temporary Migration, Parliament of Australia, *Temporary Migration* (22 July 2020) 18.

¹⁷ Department of Home Affairs (Cth), *Working Holiday Maker Visa Program Report* (30 June 2019, Report BR0110) 33–4 ('WHM Report 30 June 2019').

¹⁸ Duffy, Martin and Zhao (n 13) 28.

¹⁹ *National Agricultural Workforce Strategy* (n 1) 180.

Given that ABARES data suggest that local Australian workers contribute to one-quarter of the 101,000 workers in the Australian horticulture sector,²⁰ this leaves a sizable amount — of close to 30,000 farm workers — unaccounted for by official labour statistics. There is increasing recognition that undocumented workers form the bulk of these unaccounted farm workers. The *National Agriculture Workforce Strategy* heard evidence that supported previous studies and surveys suggesting that undocumented workers constituted up to 90% of the workforce in some major horticulture production regions.²¹ Part III of this article explores the available data on the number and contribution of undocumented workers to farm labour supply.

The introduction of border restrictions in March 2020, as a result of the COVID-19 pandemic, has impacted each of these three main groups of temporary migrant labour on farms: WHMs, SWP visa holders and undocumented migrants. Although the number of undocumented workers has remained consistent since the pandemic because international border restrictions have meant that there have been no new arrivals into this cohort, by September 2020 WHMs and SWP workers had fallen by 48%.²² This article now turns to a discussion of how the COVID-19 pandemic has impacted farm labour supply in Australia.

2 *The Impact of COVID-19 on Farm Labour Supply*

International border restrictions have prevented the arrival of WHMs who have typically formed a substantial part of the harvest workforce (see Table 1 below). In 2019–20 there was a 28.6% reduction in WHM visas granted (see Table 2 below).

Border restrictions also curtailed the operation of the SWP, which has been growing steadily since its inception in 2008, peaking at 12,202 visa approvals in 2018–19.²³ There were 9,824 visa approvals in 2019–20 up to March 2020 when visa processing was suspended due to Australia's international border closure. Nonetheless, the low rate of COVID-19 in Pacific countries, coupled with acute demand for labour by Australian growers, led to a number of initiatives during the pandemic to restart travel from Pacific countries to Australia under the SWP. In August 2020, the Australian Government announced a broader agreement to resume the SWP that states can opt in to.²⁴ Different state models have since been adopted. A first group of 160 workers arrived in September 2020 to work on mango farms in the Northern Territory.²⁵ By February 2021, the Queensland Government had flown in 782 workers from Pacific countries onto farms.²⁶ These workers were allowed to

²⁰ Martin, Randall and Jackson (n 14).

²¹ *National Agricultural Workforce Strategy* (n 1) 190.

²² Department of Agriculture, Water and the Environment (Cth), *Agricultural Forecasts and Outlook: December Quarter 2020* (ABARES Agricultural Commodities Report Vol 10.3, 2020) 41.

²³ *National Agricultural Workforce Strategy* (n 1) 180.

²⁴ Marise Payne, 'Seasonal and Pacific Workers to Help Fill Labour Gaps' (Media Release, 21 August 2020) <<https://www.foreignminister.gov.au/minister/marise-payne/media-release/seasonal-and-pacific-workers-help-fill-labour-gaps>>.

²⁵ Matt Brann, 'Mango Industry Celebrates as Seasonal Workers from Vanuatu Touch Down in Darwin', *ABC News* (online, 3 September 2020) <<https://www.abc.net.au/news/rural/2020-09-03/vanuatu-workers-arrive-in-darwin-to-pick-mangoes/12621234>>.

²⁶ Claire Bickers, 'SA's Plans to Fly in Pacific Island Seasonal Workers on Hold as State Eyes Alternative Quarantine Models', *The Advertiser* (online, 24 February 2021).

quarantine on Queensland farms while being employed from day one of their stay in Australia. The South Australian Government also trialled a program allowing Pacific workers to quarantine for 14 days in a regional facility prior to undertaking farm work.²⁷

Table 1: The number of WHM visas granted in 2018–19²⁸

WHM grants	Number	% change from 2017–18
First year visa grants (417 visa)	142,805	6.4% decrease
Second year visa grants (417 visa)	37,418*	14.0% increase
First year visa grants (462 visa)	23,012	6.2% increase
Second year visa grants (462 visa)	5,801	73.7% increase
Total visa grants (417 and 462 visas)	209,036	0.7% decrease

* 33,768 of the 417 second year grants worked in agriculture, forestry and fishing.

Table 2: The number of WHM visas granted in 2019–20²⁹

WHM grants	Number	% change from 2018–19
First year visa grants (417 visa)	92,282	35.3% decrease
Second year visa grants (417 visa)	28,316	24.3% decrease
Third year visa grants (417 visa)	2,075*	N/A
First year visa grants (462 visa)	19,845	13.8% decrease
Second year visa grants (462 visa)	6,128	5.6% increase
Third year visa grants	603 [#]	N/A
Total visa grants (417 and 462 visas)	149,249	28.6% decrease

* 75.3% of the third WHM visa applicants indicated that they undertook agricultural work to acquire their eligibility.

[#] 42.1% of the third WHM visa applicants indicated that they undertook agricultural work to acquire their eligibility.

The COVID-19 pandemic has also limited the movement of undocumented farm workers across state and territory borders. As state and territory governments responded to the pandemic by introducing border restrictions and travel permits to cross borders, undocumented workers were restricted to working only in the state or territory in which they reside, instead of being able to follow the harvest trail and access farm work in any jurisdiction. Their inability to produce a driver's licence or other official documentation means they would be unlikely to access a cross-border travel permit. Although cross-border permits have not been consistently enforced across jurisdictions, it is likely that this change has impacted the movement of undocumented farm workers and contributed to labour shortages in the horticulture sector since the beginning of the pandemic.

²⁷ Stephen Wade and David Basham, 'Paringa Regional Quarantine for Seasonal Ag Workers' (Media Release, 14 March 2021) <<https://www.premier.sa.gov.au/news/media-releases/news/paringa-regional-quarantine-for-seasonal-ag-workers>>.

²⁸ *WHM Report 30 June 2019* (n 17) 7–8, 33.

²⁹ Department of Home Affairs (Cth), *Working Holiday Maker Visa Program Report* (30 June 2020, Report BR0110) 7–8.

In response to border restrictions preventing the flow of temporary migrants from overseas, federal and state governments have introduced a range of initiatives to encourage local Australian workers to undertake harvest work. However, such initiatives have had limited success, even prior to the pandemic when unemployment rose. The Seasonal Workers Incentives Trial, which aimed to channel long-term unemployed Australian workers into the horticulture sector, only resulted in 277 workers taking part in 2017–18.³⁰ It is unsurprising, then, that such initiatives continue to be largely unsuccessful in addressing the significant shortfall of overseas workers in this COVID-19 era. For example, the Relocation Assistance to Take Up a Job program, which offers Australian workers up to \$6,000 to undertake harvest work in regional areas, attracted only 148 workers in its first month of operation.³¹ In sum, the COVID-19 pandemic has significantly affected farm labour supply in Australia. This has led to the emergence of acute and escalating labour challenges on farms.

III The Profile and Prevalence of Undocumented Workers on Farms

The precise number of undocumented workers in Australia is unknown. It is difficult to gather data on the profile of undocumented workers because of their interest, and that of their employers, in not being detected. Considering uncertainty in estimated numbers, there are between 60,000 and 100,000 undocumented workers in Australia.³² The Department of Home Affairs estimated that there were 62,900 visa overstayers as at 30 June 2018.³³ In most cases, these visa holders entered Australia on a subclass 600 visitor visa and applied for an asylum visa once onshore. However, as Wright and Clibborn note, departmental estimates are problematic given the sheer number of temporary migrants without any work rights, those with restricted work rights and the mounting numbers in recent years.³⁴ Thus, official figures likely underestimate the number of undocumented workers in Australia.

Undocumented workers are identified in a range of industries such as hospitality, massage, cleaning and construction, and there is now sufficient evidence establishing the significant prevalence of undocumented workers on Australian farms.³⁵ The horticulture sector is one where undocumented workers are more likely

³⁰ Howe et al, *Durable Future Report* (n 6) 130 (Table 14.1).

³¹ Marty McCarthy and Lucy Barbour, 'Farm Labour Incentives Failing and the Result Could be Crops Left Unharvested', *ABC News* (online, 4 December 2020) <<https://www.abc.net.au/news/rural/2020-12-04/harvest-in-jeopardy-as-farm-labour-incentives-fail/12947974>>.

³² *National Agricultural Workforce Strategy* (n 1) 206. Department of Immigration and Border Protection (Cth), *BE17/172 – Visa Overstays for the Financial Year — Programme 1.2: Border Management* (Budget Estimates Hearing, Question Taken on Notice, 22 May 2017) 2. Numbers are rounded, which may result in rounding errors. See also Stephen Howells, *Report of the 2010 Review of the Migration Amendment (Employer Sanctions) Act 2007* (Report, 2011) 94.

³³ Department of Home Affairs (Cth), *Incoming Government Brief* (2019) 172.

³⁴ Chris F Wright and Stephen Clibborn, 'Back Door, Side Door, or Front Door? An Emerging De-Facto Low-Skilled Immigration Policy in Australia' (2017) 39(1) *Comparative Labor Law & Policy Journal* 165, 176.

³⁵ Marie Segrave, *Exploited and Illegal: Unlawful Migrant Workers in Australia* (Report, July 2017) <https://arts.monash.edu/_data/assets/pdf_file/0020/1532063/exploited-and-illegal_unlawful-migrant-workers-in-australia.pdf>.

to congregate because of the sector's reliance on unregulated labour hire contractors. This means growers can employ undocumented workers at arm's length and undocumented workers are more likely to find work in horticulture than in other industries because they can access this work through a contractor. Further, the sector's geographic dispersion and the remote locations of many farms means that undocumented workers are less likely to be detected in the labour market.

Rimmer and Underhill posit that the number of undocumented workers in the horticulture industry is likely to be a third of the harvest workforce.³⁶ Evidence given to a parliamentary inquiry by the Office of the Chief Trade Adviser to the Australian Government in 2016 suggested that the use of the SWP by growers had been marred by widespread reliance on 'existing cheaper sources of labour such as illegal workers'.³⁷ A discussion paper by Doyle and Howes in 2015 found that 'the use of illegal labour still seems to be widespread in the horticulture sector. Four out of five growers ... recognized that it was prevalent to at least some extent in the industry'.³⁸

A *Industry and Growers Perspectives on the Prevalence of Undocumented Workers*

There is increasing recognition by farm industry associations and individual growers that undocumented migrants form a critical part of the harvest workforce. According to these assessments, in some Australian growing regions, undocumented workers form the majority of the harvest workforce. A three-year study of labour use on Australian farms found that the horticulture industry had a 'structural reliance' on undocumented workers as a key source of farm labour.³⁹ The study, entitled, *Towards a Durable Future: Tackling Labour Challenges in the Australian Horticulture Industry* ('Durable Future Report'), interviewed growers and industry association officials who reported widespread use of undocumented workers on farms. For example, a Northern Territory grower estimated that undocumented workers comprise close to one fifth of the horticulture workforce in the greater Darwin region.⁴⁰ A Victorian industry association official suggested that 80% to 90% of the Mildura and Robinvale workforces rely on undocumented workers.⁴¹ A Wanneroo grower posited that 70% to 80% of the workforce in that region were undocumented workers, while another grower estimated that across Western Australia, at least half of the State's harvest workforce comprised of undocumented workers.⁴² The study found a problematic relationship between labour challenges facing the industry and the industry's reliance on undocumented workers.

³⁶ Malcolm Rimmer and Elsa Underhill, 'Temporary Migrant Workers in Australian Horticulture: Boosting Supply but at What Price?' in Massimo Pilati, Hina Sheikh, Francesca Sperotti and Chris Tilly (eds), *How Global Migration Changes the Workforce Diversity Equation* (Cambridge Scholars Publishing, 2015) 143, 145.

³⁷ Joint Standing Committee on Migration, *Seasonal Change: Inquiry into the Seasonal Worker Programme* (Report, May 2016) 98 [8.38].

³⁸ Jesse Doyle and Stephen Howes, *Australia's Seasonal Worker Program: Demand-Side Constraints and Suggested Reforms* (Discussion Paper, 2015) 25.

³⁹ Howe et al, *Durable Future Report* (n 6) 36.

⁴⁰ Ibid 39.

⁴¹ Ibid.

⁴² Ibid.

For example, a Wanneroo grower who was interviewed for the purposes of the study said:

If we take them [undocumented workers] out, I don't know what the level beyond crisis is, but that's where we're at. The reason I've cut back [on production] is because I cannot get enough legitimate workers and I cannot afford to take the risk of dealing with the people that are questionable ...⁴³

A Robinvale grower in the same study reported a similar view, stating that 'there's a percentage of workers around here that don't have work visas. ... That's a real worry for us because if you take them away, I don't know what Plan B is'.⁴⁴

In the most disturbing quote from the study, a labour supply contractor in Victoria observed:

In some areas in peak periods, and it still happens to this very day, I could walk into a paddock or into a street, and I could put on a blindfold and have a shotgun and twirl myself around, aim the shotgun in any direction, fire it, and there's a good likelihood that I'll hit an illegal worker ...⁴⁵

Thus, accounting for regional variations, the horticulture industry has a substantial reliance on undocumented workers. A number of growers and other stakeholders report that growers in some regions have no choice but to engage undocumented workers because of inadequate labour supply from legal sources of labour.

There is increasing acknowledgment by the horticulture industry that the employment of undocumented workers is not a marginal phenomenon. A 2019 survey performed by the Victorian Farmers Federation ('VFF') in the Sunraysia region found that 71% of growers believed they had undocumented workers working on their farm, with undocumented workers accounting for up to 28% of the total workforce in the region.⁴⁶ VFF Vice-President Emma Germano stated, '[t]he grave reality is that undocumented workers account for a large proportion of Australia's seasonal harvest workforce. Farmers cannot share information that reflects this reality for fear of reprisal from Government agencies'.⁴⁷

In 2018, a discussion paper by the National Farmers' Federation advocating for visa reform to address labour shortages acknowledged the use of undocumented workers on farms. The discussion paper observed:

While it may be naïve to suggest that no members of the sector take advantage of illegal workers because they are cheaper than legitimate labour, by-and-large, where farmers are associated with these arrangements, it's because of their chronic labour shortages and the fact that they have little alternative.⁴⁸

⁴³ Ibid 40.

⁴⁴ Ibid.

⁴⁵ Ibid 41.

⁴⁶ Victorian Farmers Federation, *VFF Calls for Ag Visa to Legalise Workers* (7 November 2019) <https://www.vff.org.au/wp-content/uploads/2020/10/VFF-MR-_72_-VFF-calls-for-Ag-Visa-to-legalise-work.pdf>.

⁴⁷ Ibid.

⁴⁸ National Farmers' Federation, *Ag Visa Discussion Paper* (2018) (copy on file with author).

In 2020, a submission made by MADEC Australia⁴⁹ to a government inquiry stated that a ‘significant cohort’⁵⁰ of the horticulture workforce consists of undocumented workers and that within the sector there are ‘employment models that incorporate use of illegal labour as a standard practice’.⁵¹ The submission also notes the long-term presence of undocumented workers in the sector and their value to growers:

Often illegal workers have lived in the community for many years, are productive and reliable and it is easy to see why growers value their contribution. However, it cannot be denied that growers are also gaining a distinct financial advantage as illegal workers have substantially lower overheads due to not receiving super or workcover services. They are also unlikely to complain about working conditions or piece rates as they have a fear of being uncovered if they report issues to government agencies.⁵²

In summary, undocumented workers form a critical part of the workforce on Australian farms, though their contribution is usually masked and not acknowledged.

B *A Review of Government Efforts to Detect Undocumented Workers*

Detection of undocumented workers has been largely ineffective and failed to address the horticulture sector’s structural reliance on undocumented workers. It is important to acknowledge that these detection efforts have failed for over two decades. As far back as 1999, the Department of Immigration and Multicultural Affairs found substantial numbers of undocumented workers and recommended increased penalties on employers. At the time, horticulture industry associations opposed this on the basis that ‘it was not always possible to attract sufficient legal workers during the harvest’.⁵³ Since then, despite the deployment of considerable resources and the development of a special taskforce in 2015,⁵⁴ these detection efforts have been so futile that undocumented workers have grown to become a key and significant part of the sector’s workforce.

Undocumented workers have an incentive to remain invisible to authorities because they risk deportation if detected. Evidence in the *Durable Future Report* suggests that undocumented workers tend to be located in more isolated areas and keep to themselves. As one local representative in the Wide Bay–Burnett region of Queensland reported, ‘[a] lot of people in town may not even see them. They sleep. They work. They sleep. They go back to Bundy and get supplies, they come back. Yeah, they’re very quiet.’⁵⁵ A representative of the Fair Work Ombudsman reported,

⁴⁹ MADEC is a not-for-profit organisation that has been operating within the horticulture sector for over 50 years and has a contract with the Department of Education, Skills and Employment to operate the Harvest Trail Information Service. See MADEC Australia Pty Ltd, Submission No 70 to Joint Standing Committee on Migration (Cth), *Inquiry into the Working Holiday Maker Program* (2019).

⁵⁰ Ibid 4.

⁵¹ Ibid 5.

⁵² Ibid.

⁵³ Department of Immigration and Multicultural Affairs (Cth), *Review of Illegal Workers in Australia: Improving Immigration Compliance in the Workplace* (Report, 1999) 27.

⁵⁴ See below nn 63–6 and accompanying text.

⁵⁵ Howe et al, *Durable Future Report* (n 6) 40.

‘[w]e hear about all these illegal workers, but [when we visit farms] we just don’t see them.’⁵⁶

It seems that undocumented workers are adept at avoiding detection and have sophisticated, well-developed strategies in the event of a Border Force raid. In the *Durable Future Report*, an officer from the Stanthorpe police force in Queensland observed that during a raid, ‘it’s like mice abandoning the ship’,⁵⁷ with a former undocumented worker describing the need to respond quickly upon the arrival of enforcement authorities: ‘Someone just shouts, “Immigration!” And that’s it, forget about your harvesting, and everything, and your lunchbox and fssht! Whoa! ... I’ve been running for almost 5 years’.⁵⁸

This is consistent with media reports indicating that undocumented workers shout code words to alert other undocumented workers in the same row that a raid is occurring,⁵⁹ and also with the opening anecdote of a 2006 Senate Committee inquiry report into harvest labour:

[Senate Committee] members touring an isolated farm just north of Euston on the Murray River, suddenly came across a team of grape pickers hard at work. Taking fright at our unexpected appearance, they fled down the vine rows toward the other end of the field. The committee had been mistaken for immigration officers conducting a raid on illegal workers.⁶⁰

Undocumented workers are also difficult to detect because of their tendency to be housed in private dwellings (usually share houses), often owned or leased by their contractor. As one labour hire contractor reported in the *Durable Future Report*, ‘[t]hey [undocumented workers] just hide. Like let’s say, you rent a house and something like that and you can hide easy.’⁶¹ A former undocumented worker described how contractors are vigilant in ensuring undocumented workers do not make local connections and remain concealed, describing how ‘the contractor will just sleep in the car outside [the houses of undocumented workers] and watch them. So no-one walks outside.’⁶²

The Australian Border Force, which is the enforcement agency housed within the Department of Home Affairs, is charged with the responsibility of detecting undocumented workers. In 2015, the Australian Government established a specialist multi-agency taskforce, known as Taskforce Cadena, to target and disrupt the organisers of visa fraud, illegal work and the exploitation of foreign workers.⁶³ There is only one mention of Taskforce Cadena’s work in the 2017–18 and 2018–19 annual

⁵⁶ Ibid.

⁵⁷ Ibid.

⁵⁸ Ibid.

⁵⁹ Emma Field, ‘Illegal Malaysian Farm Labourer Told About “Kangaroo Call” When Starting Work’, *The Weekly Times* (online, 6 August 2015) <<https://www.weeklytimesnow.com.au/news/national/illegal-malaysian-farm-labourer-told-about-kangaroo-call-when-starting-work/news-story/eb02f7037082803af6ced386775540d6>>.

⁶⁰ Senate Standing Committee on Employment, Workplace Relations and Education, Parliament of Australia, *Perspectives on the Future of the Harvest Labour Force* (Report, October 2006) vii.

⁶¹ Howe et al, *Durable Future Report* (n 6) 40.

⁶² Ibid.

⁶³ Australian Border Force, Australian Government, ‘Taskforce Cadena’ (Web Page, 16 November 2021) <<https://www.abf.gov.au/about-us/taskforces/taskforce-cadena>>.

reports of the Department of Home Affairs.⁶⁴ The former report states that the Taskforce has completed 17 operations and executed 24 warrants to investigate visa fraud, illegal work and the exploitation of foreign workers.⁶⁵ The 2018–19 report refers to ‘the execution of a combination of 50 *Migration Act* and *Crimes Act* warrants, issuing of 40 Illegal Worker Warning Notices and the referral of 25 matters to partner agencies including allegations of human trafficking’.⁶⁶ Additionally, the annual reports of the Department of Home Affairs suggest a steady decline in compliance activity targeting the location of undocumented workers and employers who employ these workers, despite the number of such workers and employers being at record levels.⁶⁷ It is highly unlikely that this reported level of enforcement activity is able to address the scale of undocumented work in the horticulture labour market.

IV A Proposed Model for Status Resolution

There are different ways in which a one-off status resolution process can be introduced. International examples of regularisation programs demonstrate that determining eligibility can be a challenge. A study of regularisation programs in Europe found that criteria for eligibility generally included length of residence, employment or possibility of future employment (for example sponsorship by an employer), humanitarian concerns, and in some cases integration into the local society and academic or professional qualifications.⁶⁸ This study noted that

between 1973 and 2008, 68 programs were implemented in Europe; a few targeted multiple groups of people, and over half were based on labor regulation [sic]. Of those people granted regularization during this period, 87 percent were unauthorized labor migrants.⁶⁹

The Centre on Migration, Policy and Society at the University of Oxford produced a study on regularisation programs for undocumented migrants in nine European Union countries and the United States (‘US’).⁷⁰ The study identified a number of factors that contribute to implementation challenges including: a lack of publicity; having overly strict requirements that limit migrant participation; application fraud; lack of administrative preparation; and the reversion of legalised immigrants to undocumented status.⁷¹ A United Nations policy brief on the impact of COVID-19 on international migration proposes that countries should explore ‘various models of regularisation pathways for migrants in irregular situations’ as part of their

⁶⁴ Department of Home Affairs (Cth), *Annual Report 2017–18* (2018) 29; Department of Home Affairs (Cth), *Annual Report 2018–19* (2019) 22.

⁶⁵ Department of Home Affairs (Cth), *Annual Report 2017–18* (n 64) 29.

⁶⁶ Department of Home Affairs (Cth), *Annual Report 2018–19* (n 64) 22.

⁶⁷ Abul Rizvi, ‘Peter Dutton’s Legacy — Australia’s Biggest Labour Trafficking Scam’, *Independent Australia* (online, 4 April 2021) <<https://independentaustalia.net/politics/politics-display/peter-duttons-legacy--australias-biggest-labour-trafficking-scam,14954>>.

⁶⁸ Kate Brick, ‘Regularizations in the European Union: The Contentious Policy Tool’ (Migration Policy Institute, December 2011).

⁶⁹ *Ibid* 7.

⁷⁰ Amanda Levinson, *The Regularisation of Unauthorized Migrants: Literature Survey and Country Case Studies* (Centre on Migration, Policy and Society, University of Oxford, 2005) 6.

⁷¹ *Ibid* 6–7.

responses to the COVID-19 pandemic.⁷² It states that ‘this crisis is an opportunity for countries to “recover better” through socioeconomic inclusion and decent work for people on the move’.⁷³ Drawing on this international research examining various models of regularisation programs, this part develops a proposed model that is intended for undocumented migrants in the horticulture sector, recognising that not all undocumented migrants are in this sector and that the proposed model can be scaled up to operate on a more general basis across all sectors. However, given the likely political constraints in introducing status resolution, the proposed model is both modest and one that can be introduced using the existing visa framework.

A *Modification to the Temporary Activity Visa (Subclass 408)*

In response to the COVID-19 pandemic, the Australian Government has already established a visa framework that could be adapted to enable undocumented workers to regularise their status, by way of the Australian Government Endorsed Event (‘AGEE’) Stream of the Temporary Activity (subclass 408) visa. Status regularisation could be achieved by extending the eligibility criteria under the subclass 408 visa to permit undocumented workers to validly apply for the visa. In other words, it can be achieved with minimal amendments to the *Migration Regulations 1994* (Cth), and within the context of existing frameworks established to respond to the COVID-19 crisis.

In February 2020, in response to international border closures and persistent labour shortages in ‘critical sectors’ including horticulture, the Australian Government extended the AGEE Stream of the subclass 408 visa to applicants engaged in work related to a ‘COVID-19 Event’ — that is, working in a critical sector related to COVID-19 rebuilding efforts and either unable to leave Australia or apply for another visa due to the pandemic. Eligibility for the subclass 408 visa was extended to applicants who did not currently hold a valid visa, but had held a visa in the past 28 days.⁷⁴ As a matter of policy, the Department of Home Affairs undertook to grant AGEE-Stream subclass 408 visas for a period of 12 months, although the maximum grant period for the visa is four years.⁷⁵

Status regularisation may be extended to undocumented workers with a demonstrated history of employment in the horticulture industry by including the subclass 408 visa among the classes of visa for which applicants may validly apply in Australia, despite having previously had a visa application refused or a visa cancelled.⁷⁶ This could be supported by an amendment to the eligibility requirements

⁷² United Nations, *Policy Brief: COVID-19 and People on the Move* (June 2020) 3.

⁷³ *Ibid* 11.

⁷⁴ This was achieved by the *Migration (LIN 20/122: COVID-19 Pandemic event for Subclass 408 (Temporary Activity) Visa and Visa Application Charge for Temporary Activity (Class GG) Visa) Instrument 2020*, introduced with effect from 4 April 2020.

⁷⁵ Department of Home Affairs (Cth), *Temporary Activity Visa (Subclass 408): Australian Government Endorsed Events (COVID-19 Pandemic Event)* (Web Page, 1 November 2021) <<https://immi.homeaffairs.gov.au/visas/getting-a-visa/visa-listing/temporary-activity-408/australian-government-endorsed-events-covid-19>>; Department of Home Affairs (Cth), *Subclass 408: Temporary Activity Visa* (Web Page, 1 July 2021) <<https://immi.homeaffairs.gov.au/visas/getting-a-visa/visa-listing/temporary-activity-408>>.

⁷⁶ *Migration Regulations 1994* (Cth) reg 2.12.

for subclass 408 in sch 2 of the *Migration Regulations 1994* (Cth) to include applicants who do not currently hold a valid visa, but are able to provide evidence of having worked in the horticulture industry for a period of six months. Such evidence may include: letters of support (or other evidence of employment) from former or prospective employers; evidence of continued residence in a growing region; letters of endorsement from labour unions or industry associations; labour market data showing labour shortages in a particular growing region; and/or an offer of employment from a prospective employer based on past experience. A selection of this evidence could be provided with an application, although there does need to be appropriate recognition of the difficulties for undocumented migrants in sourcing evidence of their prior work given their precarious immigration status.

The Temporary Activity (subclass 408) visa is an attractive vehicle for status regularisation for a number of reasons. First, it is an existing visa that has recently been extended (through the AGEE Stream) to workers who are unable to leave Australia and are contributing to economic rebuilding efforts. Second, the threshold requirements and fees involved in validly applying for the visa are minimal⁷⁷ and therefore suit the unique situation faced by undocumented workers, who may have limited evidence of former employment or finances to support a visa application. Third, the subclass offers a grant period of up to four years, making it attractive to undocumented workers as a means of regularising their status. Fourth, implementing status regularisation by *permitting* undocumented workers to validly apply for a visa (without positively *granting* them a visa) ensures the integrity of the regularisation process. That is, it will be the responsibility of undocumented workers to apply proactively for the visa and meet the eligibility criteria, in order to achieve status resolution — rather than this being done uniformly, by way of a blanket policy implemented by the Australian Government.

As part of this adjustment to the visa framework, the Australian Government should undertake to refrain from investigations of employers in the horticulture sector who have previously engaged undocumented workers either directly or through contractors. This is an important component of the reform package to ensure industry buy-in and broad support and cooperation by growers. It also appropriately acknowledges the challenges growers have faced for many years in accessing legal farm labour.

B *A Pathway to Permanency within the Proposed Framework*

A status resolution program should be accompanied by the creation of a permanent visa pathway to encourage uptake and participation. A failure to provide a genuine commitment to develop a permanent pathway will mean that the one-off status resolution will not succeed in incentivising sufficient numbers of undocumented workers to regularise their immigration status. Lessons from other countries that have introduced regularisation processes suggest that the model of status

⁷⁷ Ibid sch 1 cl 1237.

regularisation is important and that the incentive must be strong enough or very few undocumented workers will come forward.⁷⁸

Thus, the introduction of a one-off regularisation process should be accompanied by an announcement that a self-nominated permanent visa pathway will be made available to workers after two years' work in horticulture while holding a Temporary Activity (Subclass 408) visa. Existing permanent visa pathways will need to be modified to permit subclass 408 visa holders to transition to permanent residency.⁷⁹ The design of appropriate pathway requirements should be the subject of further stakeholder consultation within the coming 12 months to ensure that it is accessible to applicants and accommodates the unique conditions of the horticulture labour market.

C *Status Resolution in Historical Context*

It is important to acknowledge that the introduction of a status regularisation process via the subclass 408 visa is not as dramatic a break with past immigration policy as it may first appear. Even in the 1980s under the Fraser Government a 'Regularisation of Status Program' was introduced to enable 'lawful and unlawful non-citizens who had arrived before 1 January 1980 to apply for permanent residency by the end of the calendar year'.⁸⁰ Over a six-month period from 30 June to 31 December 1980, 14,000 individuals from more than 90 countries were accepted through the regularisation program.⁸¹ According to news reports at the time, only one applicant was rejected — an escapee from an overseas mental hospital who had been convicted of manslaughter — and the two oldest applicants under the 1980 scheme were elderly undocumented migrants who had managed to live illegally in Australia for decades before presenting themselves to authorities at the age of 95 and 73.⁸²

Even after the introduction of the 1980 scheme, there was a continuing emphasis on encouraging undocumented workers to come forward through a general status resolution policy, although it was not universally offered. In a number of farm

⁷⁸ See generally, Levinson (n 70); Brick (n 68); Sébastien Chauvin, Blanca Garcés-Mascareñas and Albert Kraler, 'Working for Legality: Employment and Migrant Regularization in Europe' (2013) 51(6) *International Migration* 118. Recently, US President Biden has sent a Bill to Congress that, if passed, would allow 'undocumented individuals to apply for temporary legal status, with the ability to apply for green cards after five years if they pass criminal and national security background checks and pay their taxes': see The White House, 'Fact Sheet: President Biden Sends Immigration Bill to Congress as Part of His Commitment to Modernize our Immigration System' (Press Release, 20 January 2021) 2.

⁷⁹ There are existing barriers that would prevent WHMs or SWP applicants from accessing the permanent visa pathway. For instance, former SWP visa holders are subject to condition 8503, which prevents them from applying for another visa while in Australia, unless the condition is waived. While SWP participants may have had condition 8503 waived to enable them to apply for subclass 408 visas, a separate waiver request would need to be made (and granted) in order for those visa holders to apply for a permanent visa in the future.

⁸⁰ Australian National Audit Office ('ANAO'), *Onshore Compliance — Visa Onstayers and Non-Citizens Working Illegally* (ANAO Audit Report No 2, 2004–05) 27 <https://www.anao.gov.au/sites/default/files/ANAO_Report_2004-2005_02.pdf>.

⁸¹ Else Kennedy, 'Amnesty Debate Brings Back Memories', *Sunraysia Daily* (online, 27 March 2021) <<https://www.sunrasyadaily.com.au/politics/4002115/amnesty-debate-brings-back-memories>>.

⁸² *Ibid.*

regions, the Department of Immigration and Border Protection (as it then was) worked with local stakeholders to implement a status dispute resolution approach that sought to identify undocumented workers in a particular location and to assist them to regularise their visa status.⁸³ This was a more successful method of encouraging undocumented workers to come forward and seek legal and visa assistance. While status resolution is still a service undertaken by the Department of Home Affairs and is listed on its website,⁸⁴ these outreach activities, which encourage undocumented workers to self-report, are now far less frequent.

D *Alternatives to Status Resolution*

This article has already identified the difficulties in detection and deportation within the Australian context in addressing the substantial presence of undocumented migrants on farms. Even with substantial resources, it is unlikely that stronger border management can address the scale of the current enforcement challenge. However, it is important to consider whether there are any other alternatives to status resolution or detection and deportation that may be efficacious. A number of inquiries have proposed a federal labour hire licensing scheme and this is a reform that the Australian Government has committed to introducing.⁸⁵ It is possible that federal labour hire licensing can make a contribution to improving labour standards in the horticulture sector and addressing some growers' reliance on undocumented migrants. It is also possible that improving the substitution effect between the WHM program and SWP may lead to better compliance with labour standards on farms.⁸⁶

As set out in this article, segmentation in the horticultural labour force has been a key factor, driving down wages and conditions across the industry. The substitution of temporary and undocumented workers in certain regions has suppressed wages to the extent that work in the horticulture sector is unattractive to most potential workers. Differential visa status, and the ability of labour agents to leverage this in an attempt to lower production costs for growers, are the key causes of the labour exploitation rife throughout the industry. Immigration reform is critical to stabilising and improving wages and conditions in the industry.

Although there are other reforms that can be undertaken to address labour shortages and exploitation, for example, labour hire licensing or regulating the WHM visa to match the worker-protective requirements in the SWP, none of these address the core challenge, which is the industry's structural reliance on undocumented workers. These reforms do not remove the existence of this cohort within the horticultural labour market.

⁸³ Howe et al, *Durable Future Report* (n 6) 41.

⁸⁴ Department of Home Affairs (Cth), *Status Resolution Service* (Web Page, 30 May 2019) <<https://immi.homeaffairs.gov.au/what-we-do/status-resolution-service>>.

⁸⁵ Australian Government, *Australian Government Response: Report of the Migrant Workers' Taskforce* (Report, March 2019) 3.

⁸⁶ Joanna Howe, Diane van den Broek, Alex Reilly and Chris F Wright, 'A Tale of Two Visas: Interrogating the Substitution Effect between Pacific Seasonal Workers and Backpackers in Addressing Horticultural Labour Supply Challenges and Worker Exploitation' (2018) 31(2) *Australian Journal of Labour Law* 209 ('A Tale of Two Visas').

Even in jurisdictions where labour hire licensing has been introduced, it is difficult for the licensing authority and compliance units to meet the scale of the enforcement challenge. For example, in Queensland, only one in nine contractors were audited and investigated in the first year that licensing was introduced.⁸⁷ Less than 1% of the total number of applications for licences were refused or given a conditional licence.⁸⁸ Only two contractors had their licence revoked and 68 had their licence suspended.⁸⁹ The paucity of these numbers — which were across all industries, although a focus was given to horticulture and poultry — points to the difficulty in labour hire licensing being the mechanism by which the horticulture sector reduces its reliance on undocumented workers.

Thus, labour hire licensing is part of the solution to endemic exploitation, but cannot address the substantial presence of undocumented migrants on farms unless there is an incentive for this cohort to come forward and regularise their status. Being on a valid visa will mean undocumented migrants will be far less likely to seek employment with unscrupulous contractors.

E *Common Objections to Status Resolution*

1 *Status Resolution Sets a Bad Precedent*

Senator Michaela Cash has argued against status resolution on the basis that ‘an amnesty would send a dangerous message that it’s okay to flout our strong visa and migration rules — principles that this government has worked incredibly hard over a period of time to secure’.⁹⁰ Similarly, at Senate Estimates, Secretary for the Department of Home Affairs Michael Pezzullo stated:

[i]t’s a matter for government in the end because governments can change policy in terms of guidance issued by ministers under the *Migration Act*. But, as a matter of policy, it would not be our advice to change direction, because of the perverse incentive created to get to Australia, overstay your visa and go to ground. Periodic so-called amnesties ... would create an incentive for people to get themselves smuggled into Australia, effectively on false pretences ... until such time as a government of the day said, ‘Amnesty time — now come forward’. You would just get recurring cycles. The policy is a matter for government, but I certainly wouldn’t be advising them to go down that path.⁹¹

This objection is based on the premise that status resolution would send a dangerous message that it is acceptable to enter Australia and become an unlawful non-citizen in the hope that it will one day lead to a temporary work visa or even permanent

⁸⁷ *National Agricultural Workforce Strategy* (n 1) 197.

⁸⁸ *Ibid.*

⁸⁹ *Ibid.*

⁹⁰ Senate Legal and Constitutional Affairs Legislation Committee, Parliament of Australia, 2020–21 *Budget Estimates*, 19 October 2020, 134.

⁹¹ Senate Legal and Constitutional Affairs Legislation Committee, Parliament of Australia, 2020–21 *Additional Estimates*, 22 March 2021, 190–1.

residency. Although this objection appears reasonable, it has a number of weaknesses.

First, there is evidence that, in some instances, migrant workers are victims themselves, having been forced into undocumented work through a complex network of offshore and onshore labour hire contractors and migration agents who have a business model of recruiting overseas workers on visas without work rights such as tourist visas. Howells' landmark report of unauthorised work in Australia emphasised the role of offshore agents who supply tourist visas that do not permit work:

There are many people who come to Australia on a tourist visa ... but who work to support their stay ... This method of gaining access to the labour market in Australia by non-citizens has proved reasonably successful and so it becomes attractive for organisers to arrange for tourist visas and passage to Australia and then to arrange work and some form of accommodation. ...

A person then meets them on arrival and takes them to a workplace. They may not actually meet the employer, rather they perform work and they are 'paid' by the intermediary. They may move from one workplace to another.⁹²

Further, the Durable Future Report

case studies suggest that Howells' depiction of offshore networks producing an undocumented workforce is an apt description of how some undocumented workers arrive in the Australian horticulture industry. It appears that organised crime does have a role in misleading workers in their home countries and enticing them into significant debt to fund an all-inclusive package involving a visa, flights, pre-arranged accommodation and employment.⁹³

Second, upon reflection, it makes little sense that a one-off regularisation process would draw a wave of new undocumented workers given that Australia is an island nation and every new arrival by plane is forced to present to go through border security. The situation in Australia is not analogous to the US–Mexico border, for example. Moreover, if a one-off regularisation process is introduced, as proposed by the *National Agriculture Workforce Strategy*,⁹⁴ it will be largely justified on the basis of the COVID-19 pandemic and the public health imperative to vaccinate all people residing in Australia. The pandemic is clearly an exceptional, unprecedented circumstance. If status resolution is introduced, it will have been more than 41 years since the last status resolution policy. It seems illogical that someone would move to a country in anticipation of living underground and working in exploitative jobs, in hope that they will be able to commence their pathway to permanency four decades later.

Third, evidence from the introduction of status resolution arrangements abroad suggests that the precedent-setting risk of this policy can be mitigated through the design of the regularisation process and accompanying reforms. For example, in Orrenius and Zavodny's working paper on the consequences of status regularisation on undocumented migrants, the authors conclude:

⁹² Howells (n 32) 55–6 [122]–[123].

⁹³ Howe et al, *Durable Future Report* (n 6) 42, citing *ibid* 55–6 [122]–[123].

⁹⁴ *National Agricultural Workforce Strategy* (n 1) 190.

An amnesty is most likely to succeed if accompanied by a guest worker program that allows low-skilled workers to legally enter the US and either gives such workers sufficient incentives to return to their home countries or provides them with a legal way to remain permanently in the US. A successful amnesty must also incorporate workplace enforcement to eliminate job opportunities for undocumented workers and stop the cycle of illegal immigration.⁹⁵

Exemplifying this approach, the *National Agriculture Workforce Strategy* recommended that a one-off regularisation process be introduced alongside the expansion and improvement of the SWP and WHM programs, stronger border enforcement and national labour hire licensing.⁹⁶ Cumulatively, this suite of reforms will both increase and improve labour supply by creating legal pathways for temporary migrants into the Australian horticulture sector and better enforcement of immigration rules and labour standards.

2 *Tying Undocumented Migrants to Farm Work Will Create Precarity*

A second objection against the model for one-off status resolution proposed in this article is that it will tie undocumented migrants to working in the horticulture industry in order to create a pathway from a temporary work visa to permanent residency. It is certainly true that migration frameworks that link the performance of work to a migration outcome have the potential to produce vulnerability among temporary migrants.

The WHM visa extension model is emblematic of this problem. In the WHM program, visa holders are given either a second- or third-year extension on their visa if they have performed specified work in a particular industry for a set period of time. Most WHMs earn a second year on their visa after completing 88 days on a farm. This has been found to produce significant vulnerability in WHMs doing farm work.⁹⁷

The SWP is also a tied visa. It ties seasonal workers from Pacific countries to the agriculture industry. However, this visa is regulated more extensively, involves trade unions in a worker-induction process and has a rigorous pre-approval process for growers seeking to access seasonal workers. This visa is also subject to auditing and has requirements that growers be responsible for worker induction, pastoral care and accommodation. Unlike the WHM program, which has been beset by problems of worker exploitation, the SWP has proved a far better model for ensuring temporary migrants are less vulnerable to underpayment and

⁹⁵ Pia M Orrenius and Madeline Zavodny, *What are the Consequences of an Amnesty for Undocumented Immigrants?* (Federal Reserve Bank of Atlanta, Working Paper No 2004-10, 2004) 2.

⁹⁶ *National Agricultural Workforce Strategy* (n 1).

⁹⁷ Fair Work Ombudsman, *Inquiry into the Wages and Conditions of People Working under the 417 Working Holiday Visa Program* (October 2016) 25; Sarah Martin, '#88daysaslave: Backpackers Share Stories of Farm Work Exploitation', *The Guardian* (online, 26 September 2019) <<https://www.theguardian.com/australia-news/2019/sep/26/88daysaslave-backpackers-share-stories-of-farm-work-exploitation>>; Alison Rahill, 'How 88 Days Can Turn Into a Nightmare of Debt and Despair for Backpackers', *ABC News* (online, 18 July 2017) <<https://www.abc.net.au/news/2017-07-18/how-88-days-can-turn-into-a-nightmare-of-debt-and-despair/8716564>>.

mistreatment,⁹⁸ although it must be acknowledged that there are still ongoing challenges with enforcing labour standards for Pacific workers in this visa program.

The contrast between these two models demonstrates that the mere fact of a tie between the performance of work and a migration outcome is not inherently the problem, although it has the potential to create vulnerability. Where temporary migrants are on tied visas, there needs to be robust oversight of where these workers are employed and enforcement of their rights under the law.

The status regularisation model proposed in this article is to rely on the subclass 408 visa, which ties undocumented migrants to essential industries and requires evidence of six months' employment in horticulture. The subclass 408 visa is not, strictly speaking, an employer-sponsored or nominated visa category; therefore, the usual issues of employer dependence and associated possibility of exploitation do not arise.⁹⁹ While applicants under the status regularisation pathway will be required to demonstrate six months' work experience in horticulture, the visa conditions do not compel holders to remain with the same employer. Condition 8107, to which the visa would be subject, would require visa holders broadly to maintain their eligibility for the visa through ongoing employment in the horticulture sector. However, once their visas status was secured, visa holders would have the necessary bargaining power and mobility to leave employers offering sub-standard conditions.

Nonetheless, it is important that ongoing work is done by government, industry and other stakeholders to implement the key recommendations of the *National Agriculture Workforce Strategy* that seek to address the vulnerability of temporary migrants employed in the horticulture sector. In addition to a one-off status regularisation process, the Strategy proposed: national labour hire licensing; further regulation of the WHM to minimise exploitation and to mirror worker-protective elements in the SWP; and a national portal for advertising job vacancies. These reforms will be critical to ensuring subclass 408 visa holders and other categories of temporary migrants are not exploited on farms.

V Conclusion

The horticulture sector faces an urgent and immediate labour crisis that requires government action. A key dimension of this crisis is the inability of undocumented workers to respond swiftly to job vacancies and their susceptibility to exploitation. The introduction of a status regularisation process has the potential to address both challenges.

In its ground-breaking inquiry, the *National Agriculture Workforce Strategy* has strongly recommended that the Government introduce a one-off regularisation

⁹⁸ Howe et al, 'A Tale of Two Visas' (n 86); Beth Orton, 'Working Holiday Makers Not a Long-Term Solution for Horticulture', *Devpolicy Blog* (Blog Post, 19 August 2020) <<https://devpolicy.org/working-holiday-makers-not-a-long-term-solution-for-horticulture-20200819/>>.

⁹⁹ Joanna Howe, 'Contesting the Demand-Driven Orthodoxy: An Assessment of the Australian Regulation of Temporary Labour Migration' in Joanna Howe and Rosemary Owens (eds), *Temporary Labour Migration in the Global Era: The Regulatory Challenges* (Bloomsbury, 2016) 131.

process for undocumented farm workers so that they can legally work on Australian farms.¹⁰⁰ This recommendation has the potential to address both the labour shortage dimension and the exploitation dimension of the current labour crisis on Australian farms. It will have an immediate effect on labour supply by enabling undocumented workers to move freely in the industry and across state borders. It will give ethical growers access to a workforce that they have previously been unable to engage. It will also substantially reduce the vulnerability of these workers to exploitation based on their precarious immigration status and will reduce labour supply to unethical growers who continue to undercut the competition by underpaying undocumented workers. Further, by regularising the status of undocumented workers, this will enable government, industry and unions to develop broad agreement on how other visa programs, such as the SWP, can be expanded and improved to better meet growers' labour needs.

This article has sought to build on the strong recommendation in the *National Agriculture Workforce Strategy* and examine how a one-off regularisation process can be implemented in practice. It is essential that the model introduced is one that appropriately incentivises undocumented workers to come forward to regularise their status. That is why this article has proposed a model that involves a four-year temporary work visa tied to the horticulture sector, coupled with a commitment to develop a pathway to permanent residency after two years. The international literature discussed in this article demonstrates that there are myriad implementation challenges for regularisation programs and it is important that these are addressed in the design of the proposed one-off status resolution process.

There has been some concern that the introduction of a status resolution process rewards growers who have knowingly exploited these workers in the past. This article contends that rather than 'rewarding' unscrupulous growers, regularising the status of undocumented workers will instead raise prevailing employment standards in the industry and make it harder for those unscrupulous labour hire operators and growers to derive a competitive advantage by exploiting vulnerable undocumented workers. At the same time, a status resolution process will provide all growers in the industry with access to a larger formal labour pool.

¹⁰⁰ *National Agricultural Workforce Strategy* (n 1) 190.


Keeping the (Good) Faith: Implications of Emerging Technologies for Consumer Insurance Contracts

Zofia Bednarz^{*} and Kayleen Manwaring[†]

Abstract

Developments in tools powered by Artificial Intelligence ('AI') and Big Data — both existing and emerging — are predicted to have a revolutionary effect on the insurance industry in the near future. These technological advances have begun materialising at challenging times for the insurance industry in Australia. Various inquiries, including the Royal Commission into Misconduct in the Banking, Superannuation and Financial Services Industry, have uncovered evidence of insurers' unethical, and often unlawful, practices, which adversely affect consumers. The focus of this article is harm that may be caused to consumers arising out of use of AI- and Big Data-powered analytics in terms of discrimination, exclusion, and unfair prices. We analyse insurance-specific rules currently in place, including recent reforms. We focus on anti-discrimination laws, insureds' duty of disclosure, and insurers' obligations (including the duty of utmost good faith), and consider whether they sufficiently address the potential harms that using decision-making models may cause to consumers.

^{*} PhD, Research Associate, Centre for Law, Markets and Regulation ('CLMR'), University of New South Wales ('UNSW'), Sydney, Australia; Member, Allens Hub for Technology, Law and Innovation. The author acknowledges funding from Santander Financial Institute, Fundación UCEIF (Spain) and CLMR. Email: z.bednarz@unsw.edu.au; ORCID iD:  <https://orcid.org/0000-0001-6719-8101>.

[†] PhD, Senior Lecturer, UNSW Law & Justice; Member, CLMR; Allens Hub for Technology, Law and Innovation. Email: kayleen.manwaring@unsw.edu.au; ORCID iD:  <https://orcid.org/0000-0002-1970-3430>.

The authors thank attendees of the CLMR Research Symposium 2020, especially Scott Donald, John Morgan, Jeannie Paterson and Gail Pearson for valuable feedback on an earlier draft, and Maciej Rybinski (CSIRO Data61) for technological advice. However, all errors and omissions are the authors' own.

Please cite this article as:

Zofia Bednarz and Kayleen Manwaring, 'Keeping the (Good) Faith: Implications of Emerging Technologies for Consumer Insurance Contracts' (2021) 43(4) *Sydney Law Review* 455.



This work is licensed under a Creative Commons Attribution-NonDerivatives 4.0 International Licence (CC BY-ND 4.0).

As an open access journal, unmodified content is free to use with proper attribution.

Please email sydneylawreview@sydney.edu.au for permission and/or queries.

© 2021 Sydney Law Review and authors. ISSN: 1444-9528

I Introduction

The insurance industry in Australia is currently facing important challenges and opportunities. These arise from two different forces driving transformation: first, emerging technologies; and second, various inquiries into the insurance industry, and especially recommendations set out by the Royal Commission into Misconduct in the Banking, Superannuation and Financial Services Industry ('Royal Commission'). This article considers whether current law, including recent law reform, is adequate to protect consumers in relation to insurance contracts in the face of these industry disruptions.

Developments in tools powered by Artificial Intelligence ('AI') and Big Data — both existing and emerging — are predicted to soon have a revolutionary effect on the insurance industry.¹ For a data-driven industry, such as insurance, enhanced data analytics promises cost reduction, creation of new products and the potential to offer more efficient and tailored services to their customers. In this article, we focus on insurance contracts with consumers and the influence AI tools have on business-to-consumer relationships.

Studies have demonstrated that restrictive regulation may be hindering the implementation of AI tools by financial services firms.² Regulatory approaches should therefore balance two main objectives: promoting technology uptake where it can provide benefits, and addressing risks associated with its use. In this article, we focus on harm potentially caused to consumers owing to the use of AI- and Big Data-powered analytics in terms of discrimination, exclusion and unfair prices. Any legal change should only be brought about as a response to evidence about real risks or threats associated with AI and the proven inadequacy of existing rules to a new sociotechnical reality. This article provides insight into the operation of current rules

¹ Swiss Re Institute, *Data-Driven Insurance: Ready for the Next Frontier?* (Sigma No 1/2020, 29 January 2020) 2; World Economic Forum, *The Future of Financial Services: How Disruptive Innovations Are Reshaping the Way Financial Services Are Structured, Provisioned and Consumed* (Final Report, June 2015) 59–68.

² One of the key findings by the Cambridge Centre for Alternative Finance and World Economic Forum, *Transforming Paradigm: A Global AI in Financial Services Survey* (Report, 29 February 2020) 12, 79–80.

relating to underwriting of consumer insurance in the context of technological advances.

We critically examine current rules, including recent reforms, to establish whether adequate consumer protection has, or can be, achieved, or whether other interventions are needed. In this article, we focus on provisions imposing specific obligations on insurers, particularly anti-discrimination laws and provisions of the *Insurance Contracts Act 1984* (Cth) ('ICA'). Our analysis is chiefly concerned with general insurance and life insurance consumer contracts. Privacy and data protection laws are relevant to insurers' use of emerging technologies.³ However, the scope of this article does not extend to laws directly applicable to personal data collection, sharing and processing. We assume, for the purposes of this article, that insurers have already obtained consumers' data. This could happen lawfully, such as when consumers voluntarily agree to share their data with insurers in exchange for benefits such as premium reduction. We examine the extent to which the rules reviewed may constrain insurers in using consumers' data and inferences made from that data, and how insurers' access to vast amounts of consumers' data influences parties' rights and obligations under insurance contracts.

In Part II we explain both the technologies at issue and the consequent potential for consumer harm. It is rare for sociotechnical change to arise in a regulatory vacuum: just because a new or modified technology emerges, it does not mean its use is ungoverned by existing legal principles.⁴ Moves to implement new or amended legal rules 'should be justified by evidence about real threats or risks created by technological developments and/or the poor fit between existing law and new technological possibilities'.⁵ One risk often associated with use of AI decision-making is that of bias or discrimination. In Australia, insurers are mostly free to use an individual consumer's characteristics for underwriting purposes, *subject to* significant constraints imposed by anti-discrimination laws.⁶ In Part III, therefore, we analyse the application of anti-discrimination provisions when AI tools are used for extracting meaningful features from data and decision-making.

In Part IV, we examine the insured's duty of disclosure. We argue that the rationale of the insured's duty of disclosure has been significantly affected by Big Data- and AI analytics.⁷ In Part IV(A), we consider the information asymmetry between the parties and recent law reform resulting from Recommendation 4.5 of

³ Brendan McGurk, *Data Profiling and Insurance Law* (Hart Publishing, 2019) 176–202; Florent Thouvenin, Fabienne Suter, Damian George and Rolf H Weber, 'Big Data in the Insurance Industry: Leeway and Limits for Individualising Insurance Contracts' (2019) 10(2) *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 209, 227–41.

⁴ Lyria Bennett Moses, 'How to Think about Law, Regulation and Technology — Problems with "Technology" as a Regulatory Target' (2013) 5(1) *Law, Innovation and Technology* 1, 9.

⁵ Michael Guihot and Lyria Bennett Moses, *Artificial Intelligence, Robots and the Law* (LexisNexis, 2020) 21.

⁶ Jason Courtenay, 'The Insurer's Right to Choose Risk' (2017) 40(1) *Australian and New Zealand Institute of Insurance and Finance* 36, 36–7.

⁷ In the context of English law, McGurk (n 3) 123–64 argues that the insured's duty of disclosure is rendered obsolete by use of these technologies and needs to be abolished or significantly reduced.

the Royal Commission's Final Report.⁸ This replaces the consumer's duty of disclosure with a duty to take reasonable care not to make a misrepresentation to an insurer, a recommendation made to address perceived inadequacies in consumer protection under the disclosure regime. However, Recommendation 4.5 did not consider emerging technologies. Therefore, we ask in Part IV(B) whether the implementation of the Recommendation will sufficiently address the changing information balance between the parties to an insurance contract potentially brought about by Big Data and AI technologies.

Part V examines insurers' obligations towards the insured and consumers' right to know what data influenced insurers' decisions about premiums and cover. We outline specific information duties imposed on insurers, as well as consider other duties. We then turn to the utmost good faith requirement in insurance contracts and consider its potential as a safeguard against using 'black box' models for underwriting of contracts. Part VI concludes.

II Emerging Technologies in the Insurance Industry

A Definitions

The use of AI and Big Data tools has led to different ways of doing business, including in the insurance and financial services industries. Deployment of these technologies and systems in insurance is emerging, rather than mature. Sociotechnical change arising out of these technologies — that is, the new things, conduct and relationships⁹ enabled by them — have great *potential* to deliver both benefits and harms for consumers, as well as insurers.

These emerging technologies have also led to a plethora of literature on their sociotechnical attributes and affordances, as well as analysis of related concepts. A full literature review is outside the scope of this article. However, it is essential, when looking at legal problems potentially arising from sociotechnical change, that there is a good understanding of the nature of technology discussed.¹⁰ So we propose relevant definitions based on the approach of Guihot and Bennett Moses, who undertook a significant literature review relating to these technologies, in the context of legal and regulatory regimes.¹¹

⁸ *Royal Commission into Misconduct in the Banking, Superannuation and Financial Services Industry* (Final Report, 4 February 2019) vol 1, 32 ('*Royal Commission Report*').

⁹ Lyria Bennett Moses, 'Why Have a Theory of Law and Technological Change?' (2007) 8(2) *Minnesota Journal of Law, Science & Technology* 589, 594.

¹⁰ Chris Reed, 'Taking Sides on Technology Neutrality' (2007) 4(3) *SCRIPTed* 263, 282; Bert-Jaap Koops, 'Ten Dimensions of Technology Regulation: Finding Your Bearings in the Research Space of an Emerging Discipline' in Morag Goodwin, Bert-Jaap Koops and Ronald Leenes (eds), *Dimensions of Technology Regulation* (Wolf Legal Publishing 2010) 312.

¹¹ Guihot and Bennett Moses (n 5) ch 1.

1 *Artificial Intelligence and Machine Learning*

The term ‘artificial intelligence’ (AI) is over 60 years old, but still lacks consensus as to definition.¹² This is unsurprising, given the density and contested nature of the concepts involved, such as ‘intelligence’. AI exists as a technological discipline or field of study, but also as a sociotechnical concept, albeit one that generates significant controversy.¹³ Both the discipline and concept are ‘constantly evolving’.¹⁴ Therefore, this article can only provide a snapshot at one point in time.

From a technical perspective, AI encompasses a range of tools and techniques, including: Machine Learning (‘ML’); computer vision; natural language processing; speech recognition; robotics; expert systems and planning and optimisation.¹⁵ An ‘AI system’ incorporates these tools or techniques, on their own or combined, into hardware and software.¹⁶ As a sociotechnical concept, AI has been described in many ways, including as ‘systems that display intelligent behaviour by analysing their environment and taking actions — with some degree of autonomy — to achieve specific goals’ by the European High-Level Expert Group on Artificial Intelligence.¹⁷ And as ‘a collection of interrelated technologies used to solve problems and perform tasks that, when humans do them, requires thinking’ by the Australasian Council of Learned Academies.¹⁸

ML is one of the more common forms of AI used in data-rich industries such as insurance. Computing devices and software can be programmed to ‘learn’: that is, ‘modify or adapt their actions ... so that these actions get more accurate’¹⁹ over time, as measured against a ‘rational goal’.²⁰ However, one of the key limitations of this learning is that it does ‘not typically include contemplating the impact of action, reasoning about intervention, or counter-factual reasoning’.²¹ ML models also tend

¹² Ibid.

¹³ See, eg, Toby Walsh, *It’s Alive! Artificial Intelligence from the Logic Piano to Killer Robots* (La Trobe University Press, 2017) 17; House of Lords Select Committee on Artificial Intelligence, *AI in the UK: Ready, Willing and Able?* (Report of Session 2017–19, HL Paper 100, 16 April 2018) 13–14; Iria Giuffrida, Fredric Lederer and Nicolas Vermeys, ‘A Legal Perspective on the Trials and Tribulations of AI: How Artificial Intelligence, the Internet of Things, Smart Contracts, and Other Technologies Will Affect the Law’ (2018) 68(3) *Case Western Reserve Law Review* 747, 751–6; Roger Clarke, ‘What Drones Inherit from Their Ancestors’ (2014) 30(3) *Computer Law & Security Review* 247, 249.

¹⁴ Guihot and Bennett Moses (n 5) 19.

¹⁵ Toby Walsh, Neil Levy, Genevieve Bell, Anthony Elliot, James Maclaurin, Iven Mareels and Fiona Wood, *The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing* (Report for the Australian Council of Learned Academies, July 2019) 32–6.

¹⁶ Guihot and Bennett Moses (n 5) 14.

¹⁷ European Commission High-Level Expert Group on Artificial Intelligence, ‘A Definition of AI: Main Capabilities and Disciplines’, 8 April 2019, <https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341> 1. Note, however, the Expert Group’s disclaimer that its description and definition of AI ‘is a very crude oversimplification of the state of the art’.

¹⁸ Walsh et al (n 15) 14.

¹⁹ Stephen Marsland, *Machine Learning: An Algorithmic Perspective* (Chapman and Hall/CRC Press, 2nd ed, 2015) 4.

²⁰ Guihot and Bennett Moses (n 5) 23.

²¹ Ibid 15, citing Judea Pearl and Dana Mackenzie, *The Book of Why: The New Science of Cause and Effect* (Basic Books, 2018) fig 1.2.

to be empirically constructed, so outcomes are based on identification and application of *correlations* in the data, and causal reasoning is not used.²²

Many forms of ML use methods directed to detection of patterns in data. These patterns are then used in predicting future data, or in probabilistic decision-making.²³ Neural networks are used in a form of complex ML (or ‘deep learning’). Deep learning is notable for the difficulty or impossibility, even for the original programmer, to work out why a particular decision was made or outcome produced.²⁴ Success of a particular ML application can be heavily dependent on amount and quality of ‘training data’ used to teach the machine, as well as quality of methods employed and whether the assumptions underlying the initial model are updated as circumstances change.

2 *Big Data*

In this article, we adopt the following meaning for ‘Big Data’: ‘approaches, techniques and methods that involve processing data with high volume, velocity and/or variety’.²⁵ The ‘volume’ of Big Data is huge: often reported as amounting to terabytes or more,²⁶ but also expanding exponentially. ‘Velocity’ refers to data generation that is dynamic, and constantly being created or modified.²⁷ This data dynamism requires very high processing speeds, so data insights are delivered in time to be useful.²⁸ ‘Variety’ of data ‘refers to the fact that data will not all lie within a single database architecture’²⁹ and includes ‘large volumes of structured and unstructured data [held] in different formats from which insights may be drawn’.³⁰ For example, different forms of data such as images, text, audio files, video files and numbers may all be linked.³¹

The technologies discussed offer tools allowing for data analysis that are unprecedented in terms of their potential for managing large quantities of data and uncovering new correlations and trends difficult or impossible for humans to discover. Current uses of deep learning techniques have brought about a paradigm

²² Roger Clarke, ‘Why the World Wants Controls over Artificial Intelligence’ (2019) 35(4) *Computer Law and Security Review* 423, 428, Table 2. See also Kalev Leetaru, ‘A Reminder that Machine Learning is about Correlations not Causation’ *Forbes Online* (15 January 2019) <<https://www.forbes.com/sites/kalevleetaru/2019/01/15/a-reminder-that-machine-learning-is-about-correlations-not-causation/?sh=5f2b93d66161>>.

²³ Kevin P Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press, 1st ed, 2012) 1.

²⁴ Jenna Burrell, ‘How the Machine “Thinks”’: Understanding Opacity in Machine Learning Algorithms’ (2016) 3(1) *Big Data & Society* 1, 3–5; Will Knight, ‘The Dark Secret at the Heart of AI’ (2017) 120(3) *MIT Technology Review* 54, 56.

²⁵ Guihot and Bennett Moses (n 5) 9, citing Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (Sage Publication Ltd, 2014) 68.

²⁶ Rob Kitchin and Gavin McArdle, ‘What Makes Big Data, Big Data? Exploring the Ontological Characteristics of 26 Datasets’ (2016) (January–June) *Big Data and Society* 1, 6.

²⁷ Guihot and Bennett Moses (n 5) 76.

²⁸ *Ibid* 9.

²⁹ *Ibid*.

³⁰ *Ibid*.

³¹ Kitchin (n 25) 77.

shift: many models now — as opposed to traditional, statistical ML — work with unstructured data. The models themselves discover patterns in data and choose how to extract meaningful features from it. This new generation of ML models can process high volumes and variety of data to produce a wide range of inferences about individuals' likely behaviour and appetite for risk-taking.

3 *Opacity and Explainability*

The 'opacity' (or lack of transparency) of many AI and Big Data processes has attracted significant attention. This attention arises particularly in contexts where those processes are used to make (or help to make) decisions resulting in social consequences, such as a decision to grant insurance or allow an insurance claim. A seminal article by Berkeley computer scientist and sociologist Burrell³² outlines the most important types of opacity seen in algorithms and ML used for this purpose:

1. an opacity (corporate opacity) resulting from deliberate corporate (and state) secrecy, for reasons such as protecting trade secrets, limiting 'gaming', and avoiding scrutiny and/or regulation of dubious activities;³³
2. 'technical illiteracy', as most people do not have specialist skills required to read code and understand algorithmic design; and
3. opacity due to complexity arising out of:
 - (a) multi-component systems, for example, with voluminous code, large engineering teams and/or many interlinkages between modules; and
 - (b) interplay between large datasets and the way the model processes data (dimensional opacity).³⁴

This last category, 3(b), is distinctive to ML and needs further explanation. Some powerful ML models tend to 'possess a degree of unavoidable complexity' that may not be able to be resolved to a humanly comprehensible explanation, even by the designers.³⁵ This lack of interpretability is due to the 'high dimensionality',³⁶ of ML, in particular:

- the use of Big Data; that is, 'billions or trillions of data examples, and thousands or tens of thousands' of data properties or 'features';³⁷
- the nature of an ML model's mechanism in handling large numbers of heterogeneous features;

³² Burrell (n 24) 3–5. As of 11 December 2021, this article had been cited over 1200 times, according to Google Scholar <<https://scholar.google.com/citations?user=Cp9FkPYAAAAJ&hl=en>>.

³³ See Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press, 2015).

³⁴ For example, the Deep Patient ML system was able to come to accurate predictions of 'the onset of psychiatric disorders like schizophrenia' in hospital patients, but the developers have admitted they do not understand how it arrives at its predictions: Knight (n 24) 57.

³⁵ Burrell (n 24) 5.

³⁶ Ibid 2.

³⁷ Ibid 5.

- the use of mathematical optimisation techniques to deal with resource constraints; and
- logic of the ML model changing as it ‘learns’.³⁸

Three examples of this type of complexity follow.

First, emergent behaviour unable to be predicted by developers may be observed, arising from interplay between the dataset and changing logic and parameters of the model. Second,

[w]ith greater computational resources, and many terabytes of data to mine (now often collected opportunistically from the digital traces of users’ activities), the number of possible features to include ... rapidly grows way beyond what can be easily grasped by a reasoning human.³⁹

This exponential increase in features not only makes reasoning about algorithms more difficult, but can also lead to ‘subtl[e] and imperceptibl[e]’ shifts in decisions.⁴⁰ Third, ‘weighted’ inputs may not map to real-world features intelligible to humans,⁴¹ or may not be able to be mapped directly to an outcome because mathematical manipulation of model dimensions is needed to manage computing constraints.⁴²

This high-dimensionality type of complexity is particularly true of deep learning models, typically (although not always)⁴³ deep neural networks.⁴⁴

If you had a very small neural network, you might be able to understand it. ... But once it becomes very large, and it has thousands of units per layer and maybe hundreds of layers, then it becomes quite un-understandable.⁴⁵

A practical, somewhat simplified example, of a particular ML application can assist in understanding both corporate opacity and dimensional opacity. For example, in 2017 researchers employed by Sears department store company designed a model allowing someone with access to consumer shopping history (for example, through a customer loyalty scheme) to create abstract numerical profiles (called ‘vector representations’) of customers, in order to provide targeted product recommendations.⁴⁶ In this model, each customer’s profile can be described with a sequence of numeric values (such as a 200-position vector of continuous values), but these cannot be individually mapped to real-world characteristics. Vector representations like these can be used as additional input features in decision-making models, for example, for the purpose of insurance underwriting. Data brokers are

³⁸ Ibid 5.

³⁹ Ibid 9. See also Pedro Domingos, ‘A Few Useful Things to Know about Machine Learning’ (2012) 55(10) *Communications of the ACM* 78, 82: ‘[i]ntuition [f]ails in [h]igh [d]imensions’.

⁴⁰ Burrell (n 24) 9.

⁴¹ Ibid.

⁴² Ibid.

⁴³ Michael Legg and Felicity Bell, *Artificial Intelligence and the Legal Profession* (Hart Publishing, 2020) 33.

⁴⁴ A ‘deep’ neural network is a neural network of more than one hidden layer: Jerry Kaplan, *Artificial Intelligence: What Everyone Needs to Know* (Oxford University Press, 2016) 34.

⁴⁵ Knight (n 24) 60, quoting Professor Tommi Jaakola, MIT Computer Science and Artificial Intelligence Laboratory. See also Guihot and Bennett Moses (n 5) 154.

⁴⁶ Bibek Behera, Manoj Joshi, Abhilash KK and Mohammad Ansari Ismail, ‘Distributed Vector Representation of Shopping Items, the Customer and Shopping Cart to Build a Three Fold Recommendation System’ (2017) *arXiv:1705.06338 [cs.IR]*.

also likely to create and on-sell these profiles to one or more third parties. To protect their business model, creators will have no incentive to disclose any proprietary information to third parties beyond the minimum needed for integration. So third parties (such as insurers) are likely to be given a ‘black box’ to integrate into their decision-making models. Even where it might be possible to trace the impact of individual dimensions of these vectors on the model’s output (regarding their insurance risk profile) they still cannot be traced to any real-world properties of a person. However, latently, they may encode any properties of this person. Therefore, neither the insurer nor an individual insured can know, or demonstrate, what properties (for example, race, gender, religion) are encoded in an individual’s vector representation.

Understandably, due to these different forms of opacity, there have been substantial calls for ‘explainability’ of AI processes. Explainability ‘refers to any technique that helps the user or developer of ML models understand why models behave the way they do’.⁴⁷ There are different types of explainability, including ‘global explainability’, which ‘attempts to understand the high-level concepts and reasoning used by a model’,⁴⁸ and ‘local explainability’, which ‘aims to explain the model’s behaviour for a specific input’.⁴⁹

Local explainability is most useful for a consumer seeking to understand a decision.⁵⁰ Common examples of local explainability techniques used in ML include:⁵¹

- feature importance scores (for example, 60% of the decision was based on age, which had a positive correlation with likely number of claims);
- counterfactual explanations (for example, had your annual mileage been lower by 10,000km, your premium would have been 50% cheaper);⁵² and
- identification of influential training data points (for example, past claimants A, D and E had the most influence on predictions) (‘influential example identification’).

⁴⁷ Umang Bhatt, Alice Xiang, Shubham Sharma, Adrian Weller, Ankur Taly, Yunhan Jia, Joydeep Ghosh, Ruchir Puri, José MF Moura and Peter Eckersley, ‘Explainable Machine Learning in Deployment’ in *FAT* ’20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (January 2020) 648, 648.

⁴⁸ *Ibid* 649.

⁴⁹ *Ibid*. See an alternative categorisation of ‘model-centric explanations’ and ‘subject-centric explanations’: Lilian Edwards and Michael Veale, ‘Slave to the Algorithm? Why a ‘Right to Explanation’ is Probably Not the Remedy You are Looking for’ (2017) 16(1) *Duke Law & Technology Review* 18, 22.

⁵⁰ Edwards and Veale (n 49) 22.

⁵¹ Bhatt et al (n 47).

⁵² See Sandra Wachter, Brent Mittelstadt and Chris Russell, ‘Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR’ (2018) 31(2) *Harvard Journal of Law & Technology* 841, 844–6.

However, a 2020 study carried out on ‘explainable AI’ found continuing barriers to use of ML explanations by end users in theory and in practice.⁵³ General problems include the risk of spurious correlations, lack of causal explanations and the need for experts to interpret explanations.⁵⁴ For the local explainability examples set out above, difficulties include:

- feature importance analyses resulting in unexpected explanations not aligned with human intuition;
- unfeasible and suboptimal counterfactuals;
- ‘intractability’ of influential example identification for large datasets (ie no efficient model exists); and
- sensitivity of influential example identification to outliers in the data.⁵⁵

4 *Inferences*

Inference is the process of using a trained AI model to make a prediction. ML models have been shown to be capable of inferring things such as a person’s sexual orientation from their face photos,⁵⁶ or a person’s suicidal tendencies from their posts on Twitter.⁵⁷ However, a question arises as to the accuracy of such predictions. Models operate on correlations between input data and target variables, rather than confirming a causal relationship between them.⁵⁸ Consequently, where certain identified features *correlate* statistically with their risk outcome for an insurer’s model, this does not mean that risk will be correct for a specific individual.

B *Potential for Consumer Harm*

It is difficult to assess the full extent to which insurers are currently using the technologies discussed in Part II(A), particularly more complex forms of ML such as deep learning. Like many corporate entities, substantial details of technologies used by insurers are usually unknown to consumers, due to corporate secrecy practices.⁵⁹ However, there are some clear indications that insurers are increasingly using large datasets and AI and other automated tools and techniques to assist them in various business processes. These processes include: determining to whom they will offer insurance; price and conditions on which insurance is offered;⁶⁰ processing

⁵³ The study by Carnegie Mellon, Cambridge, University of Texas and IBM researchers (among others) is reported in Bhatt et al (n 47).

⁵⁴ Ibid 656.

⁵⁵ Ibid 651–5.

⁵⁶ Yilun Wang and Michal Kosinski, ‘Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation from Facial Images’ *OSF* (Research Project, updated 26 May 2020) <<https://osf.io/zn79k/>>.

⁵⁷ Bridianne O’Dea, Stephen Wan, Philip J Batterham, Alison L Caele, Cecile Paris and Helen Christensen, ‘Detecting Suicidality on Twitter’ (2015) 2(2) *Internet Interventions* 183.

⁵⁸ Anya ER Prince and Daniel Schwarcz, ‘Proxy Discrimination in the Age of Artificial Intelligence and Big Data’ (2020) 105(3) *Iowa Law Review* 1257, 1263–4.

⁵⁹ Pasquale (n 33) ch 2.

⁶⁰ Guihot and Bennett Moses (n 5) 231.

claims; fraud detection; client communications; and payments.⁶¹ For example, both the 140-year-old Zurich Insurance⁶² and neophyte Lemonade Insurance Company⁶³ have publicly announced AI use in claims handling. In 2016, some United Kingdom ('UK') insurance businesses acknowledged use of ML techniques in assessing and pricing risk for motor and home insurance,⁶⁴ and United States ('US') businesses have indicated increasing use of and interest in Big Data-powered predictive analytics for property and casualty insurance.⁶⁵ Data-rich companies outside the insurance industry (such as Woolworths and Qantas) are now branching out into insurance, offering discounts linked to the provision of more and more granular lifestyle data through rewards card schemes.⁶⁶ Potential for the use of these technologies across the industry is significant, with their promise of cost reductions, especially labour costs, efficiency growth and increase in market share.⁶⁷

Data profiling in insurance allows for personalisation of risk, and therefore individualisation of premium and cover. More precise risk assessment creates important benefits for insurers, in addition to possibly lower premiums⁶⁸ for many insureds. Personalisation of risk and pricing can incentivise consumers to adopt more prudent behaviour, when risk is under their control. Some insureds, however, will face higher premiums. Others may be considered uninsurable, and unable to change or control their risk profile.⁶⁹

Advanced Big Data analytics, and the data collection required,⁷⁰ creates opportunities for insurers to access increasingly large amounts of data on consumers. Exploitation of those opportunities will exacerbate existing information

⁶¹ Ibid 250.

⁶² Brenna Hughes Neghaiwi and John O'Donnell, 'Zurich Insurance Starts Using Robots to Decide Personal Injury Claims', *Reuters* (online, 18 May 2017) <<https://www.reuters.com/article/zurich-ins-group-claims-idUSL8N1IJ3L0>>.

⁶³ Daniel Schreiber, 'Lemonade Sets a New World Record: How A.I. Jim Broke a World Record without Breaking a Sweat', *Lemonade Blog* (Blog Post, 1 January 2017) <<https://www.lemonade.com/blog/lemonade-sets-new-world-record/>>.

⁶⁴ Financial Conduct Authority (UK), *Call for Inputs on Big Data in Retail General Insurance* (Feedback Statement FS16/5, September 2016) [2.21] <<https://www.fca.org.uk/publication/feedback/fs16-05.pdf>>.

⁶⁵ 'P/C Insurers Expanding Use of Predictive Models, Big Data across Functions', *Insurance Journal News* (Article, 3 March 2016) <<https://www.insurancejournal.com/news/national/2016/03/03/400846.htm>>.

⁶⁶ 'Woolworths Insurance: Life Insurance' (Web Page, 2020) <<https://insurance.woolworths.com.au/life-insurance.html>>; 'Qantas Insurance' (Web Page, 2021) <<https://insurance.qantas.com/>>.

⁶⁷ Organisation for Economic Co-operation and Development ('OECD'), *The Impact of Big Data and Artificial Intelligence (AI) in the Insurance Sector* (Report, 28 January 2020) 8–9.

⁶⁸ Actuaries Institute, *The Impact of Big Data on the Future of Insurance* (Green Paper, November 2016) 4.

⁶⁹ The UK car insurance market after the Court of Justice of the European Union ruling in *Association Belge des Consommateurs Test-Achats ASBL v Conseil des Ministres* (C-236/09) [2011] ECR I-00773 provides an interesting example: see Patrick Collison, 'How an EU Gender Equality Ruling Widened Inequality', *The Guardian* (online, 14 January 2017) <<https://www.theguardian.com/money/blog/2017/jan/14/eu-gender-ruling-car-insurance-inequality-worse>>.

⁷⁰ Machine learning has been described as 'very data hungry': World Economic Forum, *The New Physics of Financial Services: Understanding How Artificial Intelligence is Transforming the Financial Ecosystem* (Report, August 2018) 42.

asymmetries and power imbalances between insurers and consumers. Consequent harm will be caused if insurers abuse their increased access to information in order to refuse claims, ultimately depriving consumers of the benefit of their insurance. We discuss this issue in more detail in Part IV(A).

Use of AI tools for commercial decision-making also raises concerns regarding discrimination and bias.⁷¹ Discrimination can be direct, as when a person is treated differently because of their membership in a protected class; or indirect, when a seemingly neutral rule leads to discriminatory outcomes.⁷² In the context of algorithmic bias, the concept of ‘fairness’ is often used to refer to the need to prevent or limit indirect discrimination.⁷³ The risk of both direct and indirect discrimination does not necessarily result from the use of ML models. In fact, indirect discrimination in the provision of financial services has been a longstanding problem, with practices such as ‘redlining’ reportedly used by humans (using traditional simple algorithms) in financial services businesses long before ML and Big Data tools emerged.⁷⁴ However, in the context of use of new ML models, indirect discrimination is concerning for several reasons.⁷⁵ Considering the exponential growth of data held by organisations, technological advancements and promised increases in cost- and time-effectiveness, a growing number of people may soon be affected.⁷⁶ While the problem is not new per se, the potential for harm is arguably greater.

ML models, like simpler algorithms, may reproduce human biases and introduce new ones, potentially leading to discriminatory decision-making.⁷⁷ AI models will inevitably indirectly discriminate (not necessarily unlawfully)⁷⁸ when the protected attribute is a predictive characteristic.⁷⁹ For example, if women were

⁷¹ Centre for Data Ethics and Innovation (UK) (‘CDEI’), *Review into Bias in Algorithmic Decision-Making* (Report, November 2020) 21; Eirini Ntoutsis, Pavlos Fafalios, Ujwal Gadgiraj, Vasileios Iosifidis, Wolfgang Nejdl, Maria-Esther Vidal, Salvatore Ruggieri, Franco Turini, Symeon Papadopoulos, Emmanouil Krasanakis, Ioannis Kompatsiaris, Katharina Kinder-Kurlanda, Claudia Wagner, Fariba Karimi, Miriam Fernandez, Harith Alani, Bettina Berendt, Tina Kruegel, Christian Heinze, Klaus Broelemann, Gjergji Kasneci, Thanassis Tiropanis and Steffen Staab, ‘Bias in Data-Driven Artificial Intelligence Systems: An Introductory Survey’ [2020] 10(3) *WIREs Data Mining and Knowledge Discovery* 1, 2.

⁷² *Waters v Public Transport Corp* (1991) 173 CLR 349, 392 (Dawson and Toohey JJ). See also *Australian Iron & Steel Pty Ltd v Banovic* (1989) 168 CLR 165, 175 (Deane and Gaudron JJ); *Commonwealth Bank of Australia v Human Rights and Equal Opportunity Commission* (1997) 80 FCR 78, 96–7 (Sackville J).

⁷³ See, eg, Finn Lattimore, Simon O’Callaghan, Zoe Paleologos, Alistair Reid, Edward Santow, Holli Sargeant and Andrew Thomsen, *Using Artificial Intelligence to Make Decisions: Addressing the Problem of Algorithmic Bias* (Australian Human Rights Commission Technical Paper, November 2020).

⁷⁴ Prince and Schwarcz (n 58) 1268–9.

⁷⁵ Lattimore et al (n 73) 26, 31, 40; Prince and Schwarcz (n 58) 1267–8; CDEI (n71) 7; Frederik J Zuiderveen Borgesius, ‘Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence’ (2020) 24(10) *The International Journal of Human Rights* 1572, 1577; Ntoutsis et al (n 71) 9.

⁷⁶ CDEI (n71) 24.

⁷⁷ Ibid 21; Ntoutsis et al (n 71) 2.

⁷⁸ See Part III in this article.

⁷⁹ Lattimore et al (n 73) 32; Prince and Schwarcz (n 58) 1273.

more likely to default on a loan and sex was a protected attribute, the model would look for proxies for sex (for example, clothes purchased, Facebook likes, and Netflix selections). Consequently, using ML models for predictive analytics requires businesses to test for bias and discrimination, as it will occur if a protected attribute has predictive value.

The 2020 Australian Human Rights Commission ('AHRC') Technical Paper illustrates five ways in which algorithmic bias arises:

- (1) different base rates, when a protected group is predicted to be less profitable, because of historical and current disadvantage;⁸⁰
- (2) historical bias, 'when the data used to train an AI system no longer accurately reflects reality';⁸¹
- (3) label bias, when human bias in the recording of the target (customer profitability) is reproduced;⁸²
- (4) contextual features and underrepresentation, when patterns and trends identified in the data across individuals that influence AI model's predictions are not transferable across different demographics;⁸³ and
- (5) contextual features and inflexible models, when 'the data contains insufficient information to capture the differing behaviour of the various demographics'.⁸⁴

Algorithmic bias is often linked to human (conscious and unconscious) biases. This arises when historical data under- or over-representing certain phenomena or groups of people⁸⁵ is embedded in datasets used for training and testing of the models. In our example above, it may be that historically women were rarely granted loans, as their work was unpaid and relied on their husband's income. The model would therefore learn that loans with low default risk were only granted to men. In simple terms, for the model, discriminating against women would lead to the desired outcome; that is, maximising loans' safety and profitability for the firm. This example shows how testing for bias is necessary.

The problem also derives from lack of reliable datasets for both training and testing. Reasons for this unreliability range from technological to economic to legal. At the technological end, de-identification of data leads to complications.⁸⁶ In general terms, for a dataset to be de-identified, all traces of information potentially revealing a subject's identity have to be removed or changed.⁸⁷ The process usually involves removal of sensitive or protected attributes, such as gender-, health-, or ethnicity-related data.⁸⁸ Consequently, however, bias may become unmeasurable.

⁸⁰ Lattimore et al (n 73) 31–2.

⁸¹ Ibid 34.

⁸² Ibid 40.

⁸³ Ibid 45.

⁸⁴ Ibid 48–9.

⁸⁵ Ibid 34–45 (Scenarios 2 and 3).

⁸⁶ Christine M O'Keefe, Stephanie Otorepec, Mark Elliot, Elaine Mackey and Kieron O'Hara, 'The De-Identification Decision-Making Framework', *CSIRO Data61* (Report, 18 September 2017) 4–5.

⁸⁷ Ibid 9–10, cf *Privacy Act 1988* (Cth) s 6(1).

⁸⁸ O'Keefe et al (n 86) 18–20.

De-identification is a useful tool for privacy compliance, as information that has been de-identified is no longer considered personal information. Also, competitive strategies (and in some cases competition law) prevent businesses from sharing datasets, which means models can only be trained and tested on a firm's own historical records.

There is anecdotal evidence illustrating effects on the private sector, with examples including technology giants Amazon,⁸⁹ Google⁹⁰ and Facebook.⁹¹ Also, as noted, discrimination, especially indirect, may be difficult to observe and distinguish from treatment that is not discriminatory, as seen in the case of the Apple credit card. The card, a joint undertaking of Apple and Goldman Sachs, was launched in the US in August 2019. It was alleged that Apple's system granted men much higher credit limits than women, or even rejected women's applications for the card altogether.⁹² The issue only came to light when husband-and-wife couples applied for the card and compared outcomes. However, after investigating, the New York State Department of Financial Services ('NYSDFS') concluded in March 2021 that there had been no discrimination, and credit rates varied due to each person's different credit score, indebtedness, income, credit utilisation, missed payments, and other credit history elements.⁹³ This case however reveals an issue arising out of use of Big Data analytics: when clients receive personalised financial products, they cannot know immediately what features were taken into account, and subsequent investigation may be required.

Arguably, this has always been the case with any type of decision-making, including when done by humans or simple non-AI algorithms, as indirect discrimination is generally difficult to observe or discover. However, the problem is exacerbated by Big Data and AI technological advancements, tying back to the opacity/explainability problem discussed in Part II(A). It can be illustrated by the amount of data that could, potentially, be considered by insurers. For example, possible indirect discrimination would be even more difficult to discover if a model were basing its underwriting decision on a person's Instagram account, combined with their grocery shopping history, Google search history, websites they visited and conversations overheard by a voice assistant, such as Alexa or Siri, information collected by smart homes, and telematic data from vehicles.

⁸⁹ Jeffrey Dastin, 'Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women', *Reuters Technology News* (online, 10 October 2018) <<https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>>.

⁹⁰ Amit Datta, Michael Carl Tschantz and Anupam Datta, 'Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination' [2015] (1) *Proceedings on Privacy Enhancing Technologies* 92.

⁹¹ Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove and Aaron Rieke, 'Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes' (2019) 3(199) *Proceedings of the ACM on Human-Computer Interaction* 1.

⁹² Neil Vidgor, 'Apple Card Investigated after Gender Discrimination Complaints', *The New York Times* (online, 10 November 2019) <<https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html>>.

⁹³ New York State Department of Financial Services ('NYSDFS'), *Report on Apple Card Investigation* (Report, March 2021).

Use of genetic tests for underwriting of life insurance contracts illustrates how insurers' access to too much information can be damaging to insureds. The Financial Services Council has introduced a moratorium on genetic tests in life insurance from July 2019 to June 2024. While the moratorium is in place, industry members cannot use adverse genetic test results for purposes of underwriting of contracts, subject to certain conditions.⁹⁴ The ban on life insurance risk evaluation on the basis of adverse genetic tests results is a solution advocated by many experts,⁹⁵ and has been implemented successfully in other jurisdictions, such as the UK, Canada and some European jurisdictions.⁹⁶ This is because studies have demonstrated serious harms to consumers linked to genetic tests in the life insurance context. Insurer practices contrary to the industry's own guidelines have been uncovered,⁹⁷ such as refusing to take into account treatments undertaken by consumers lowering risk of the genetic condition occurring.⁹⁸ Also, despite an Australian Law Reform Commission recommendation, no standardised control of quality (scientific reliability, actuarial relevance and reasonableness) was made regarding the test insurers used for underwriting contracts.⁹⁹ All these issues meant consumers were reluctant to participate in medical surveys or test themselves for a genetic condition, knowing this could preclude access to life insurance.¹⁰⁰ Similarities exist between genetic information and other kinds of information about potential insureds collected through digital means and inferred with use of AI tools. The reliability of such digitally collected and inferred information, as well as its potential adverse effect on consumers, could face similar issues to genetic tests.

Although indirect discrimination is not a feature exclusively linked to ML models,¹⁰¹ there are community expectations as to the fairness of AI-based tools. If such tools are used by businesses for commercial gain,¹⁰² they should not perpetuate bias and lead to (unlawful) discriminatory outcomes. We contend that no form of unlawful discrimination, especially as it disproportionately affects already

⁹⁴ Financial Services Council, *Standard No. 11: Moratorium on Genetic Tests in Life Insurance*, 21 June 2019, 3–5.

⁹⁵ See, eg, Ainsley J Newson, Sam Ayres, Jackie Boyle, Michael T Gabbett, and Amy Nisselle, 'Human Genetics Society of Australasia Position Statement: Genetic Testing and Personal Insurance Products in Australia' (2018) 21(6) *Twin Research and Human Genetics* 533.

⁹⁶ Margaret Otłowski, Kristine Barlow-Stewart, Paul Lacaze and Jane Tiller, 'Genetic Testing and Insurance in Australia' (2019) 48(3) *Australian Journal of General Practice* 96, 98.

⁹⁷ Ibid 97.

⁹⁸ Jane Tiller, Susan Morris, Toni Rice, Krystal Barter, Moeen Riaz, Louise Keogh, Martin B. Delatycki, Margaret Otłowski and Paul Lacaze, 'Genetic Discrimination by Australian Insurance Companies: A Survey of Consumer Experiences' (2020) 28(1) *European Journal of Human Genetics* 108, 109.

⁹⁹ Otłowski et al (n 96) 98.

¹⁰⁰ Jane Tiller, Margaret Otłowski and Paul Lacaze, 'Should Australia Ban the Use of Genetic Test Results in Life Insurance?' (2017) 5 *Frontiers in Public Health* Article 330,2.

¹⁰¹ Anti-Discrimination Working Group of the Actuaries Institute ('ADWG'), 'The Australian Anti-Discrimination Acts: Information and Practical Suggestions for Actuaries' (Paper Presented to the Actuaries Institute 20/20 All-Actuaries Virtual Summit, 3–28 August 2020) 28.

¹⁰² See, eg, Ramnath Balasubramanian, Ari Libarikian and Doug McElhaney, 'Insurance 2030: The Impact of AI on the Future of Insurance', *McKinsey & Company Insurance Practice* (Article, 12 March 2021) <<https://www.mckinsey.com/industries/financial-services/our-insights/insurance-2030-the-impact-of-ai-on-the-future-of-insurance>>.

disadvantaged groups, should be accepted.¹⁰³ Therefore, insurers (alongside all businesses dealing with consumers) who use new technologies should undertake specific steps aiming to ensure the fairness of decision-making models used.¹⁰⁴

At the same time, it must be acknowledged that some form of indirect discrimination will almost always exist in any decision-making procedure,¹⁰⁵ and so it cannot be fully avoided in automated decision-making settings. The question is when it is *unreasonable* for indirect discrimination to exist, and thus when it should be considered unlawful.¹⁰⁶ With this consideration in mind, we proceed to examine relevant provisions regulating insurers' obligations and the protection offered to insureds when AI analytics are used.

III Anti-Discrimination Laws and Automated Data Profiling in Insurance Contracts

We begin by analysing anti-discrimination laws and protection available to consumers when algorithmic decision-making is used in the context of insurance contracts. Discrimination, in relation to provision of goods and services, is understood as a different, less favourable treatment, than another person would receive in the same circumstances. This less favourable treatment is due to discrimination against a certain characteristic of a person.¹⁰⁷ Not every type of different treatment is unlawful discrimination; it will only be considered as such for specific characteristics set out in the legislation. Insurers must comply with both state and federal discrimination legislation, although there are some exemptions for insurance contracts, as discussed below. At the federal level, discrimination based on the protected attributes of age,¹⁰⁸ disability,¹⁰⁹ sex (including, for example, marital status or pregnancy and breastfeeding)¹¹⁰ and race (including national or ethnic origin)¹¹¹ is forbidden. At the state level, protected attributes may also include professional or industrial activity, trade, or occupation.¹¹²

Personalisation of premiums and cover in insurance contracts results in treating insureds differently based on their characteristics, as insureds face different levels of risk. Insurers are free to choose factors on which they base the premium price (in most lines of insurance), unless anti-discrimination laws apply. However, even in the case of protected attributes, both state and federal anti-discrimination

¹⁰³ Zuiderveen Borgesius (n 75) 1574–6, see especially discussion as to gender-biased advertising by Google, which shows to men more job ads for positions with higher salary than to women: at 1575 citing Datta, Tschantz and Datta (n 90).

¹⁰⁴ Lattimore et al (n 73) 14.

¹⁰⁵ Especially in underwriting of insurance, see ADWG (n 101) 28.

¹⁰⁶ Lattimore et al (n 73) 14.

¹⁰⁷ See, eg, *Age Discrimination Act 2004* (Cth) s 14 ('*Age Discrimination Act*').

¹⁰⁸ *Ibid.*

¹⁰⁹ *Disability Discrimination Act 1992* (Cth) ('*Disability Discrimination Act*').

¹¹⁰ *Sex Discrimination Act 1984* (Cth) ('*Sex Discrimination Act*').

¹¹¹ *Racial Discrimination Act 1975* (Cth).

¹¹² *Discrimination Act 1991* (ACT), *Anti-Discrimination Act 1992* (NT), *Anti-Discrimination Act 1998* (Tas), *Equal Opportunity Act 2010* (Vic).

legislation make it lawful in *some* cases for an insurer to refuse insurance, or to discriminate on policy terms. These exceptions allow insurers to discriminate based on age, disability and gender.¹¹³ Normally, unlawful discriminatory conduct can also be permitted under a time-limited exemption.¹¹⁴ Exemptions are subject to a two-limb test. Discrimination, to be lawful, needs to be based upon actuarial or statistical data on which it is reasonable for the insurer to rely, and needs to be reasonable having regard to the matter of the data and other relevant factors (the ‘data limb’).¹¹⁵ If (and only if) such actuarial or statistical data is not available and cannot reasonably be obtained, the discrimination may be considered as reasonable having regard to any other relevant factors (the ‘no-data limb’).¹¹⁶

How, then, would an insurer’s Big Data analytics be viewed under the two limbs? Two questions should be considered. First, the possible classification of AI models under the data limb. Second, the implications of using AI models for the reasonableness of discrimination.

To answer the first question, we need to consider two distinct uses of AI models. If a model applies known actuarial or statistical data, there is no doubt it will fall under the data limb. For example, AI tools could examine potential insureds’ social media accounts, automatically searching for evidence of them engaging in high- or low-risk activities, according to known statistics. Evidence of smoking or extreme sports, such as paragliding, would indicate a higher risk to health or life, while evidence of behaviours such as regular exercise and healthy eating, would imply a lower risk. In such cases, the insurer would be able to prove they based the underwriting decision on these factors. Following the decision in *Ingram v QBE Insurance (Australia) Ltd*,¹¹⁷ the insurer must show they actually knew and applied the relevant empirical evidence at the time of making the decision. It is not sufficient to show evidence confirming the correctness of the decision regarding premium and policy terms, if this evidence was unknown to the insurer when the policy was issued.

In contrast to traditional statistical models, ML models are commonly used in data mining processes to discover meaningful patterns in data, rather than starting from a given hypothesis as to which features are predictive of the target outcome.¹¹⁸

¹¹³ At the federal level, see: *Age Discrimination Act* (n 107) s 37; *Disability Discrimination Act* (n 109) s 46; *Sex Discrimination Act* (n 110) s 41.

¹¹⁴ At the federal level: *Age Discrimination Act* (n 107) s 44; *Disability Discrimination Act* (n 109) s 55; *Sex Discrimination Act* (n 110) s 44.

¹¹⁵ Australian Human Rights Commission, *Guidelines for Providers of Insurance and Superannuation under the Disability Discrimination Act 1992 (Cth)* (Guidelines, November 2016) 9 (‘AHRC Guidelines’) considers that ‘[t]he question of whether it is reasonable for a provider to rely upon particular data involves “an objective judgment about the nature and quality of the actuarial or statistical data” in each case’, citing *QBE Travel Insurance v Bassanelli* (2004) 137 FCR 88, 95 [30] (‘Bassanelli’). The AHRC Guidelines provide examples of actuarial or statistical data that demonstrate that it is reasonable to rely upon (in relevant circumstances): underwriting manuals, local data, international studies, and relevant domestic and international insurance experience.

¹¹⁶ As per provisions listed in nn 112–13, with discrimination on the basis of sex being the main exception, as it does not allow for discrimination when actuarial or statistical data is not available: see *Sex Discrimination Act* (n 110) s 41(1).

¹¹⁷ *Ingram v QBE Insurance (Australia) Ltd* [2015] VCAT 1936 (‘Ingram’).

¹¹⁸ See Prince and Schwarcz (n 58) 1274.

As discussed in Part II(A), patterns derived by a ML model may be opaque, so that an outcome may not be easily traced to particular features of the data subject. In this scenario, the model used cannot be classified under the data limb as applying ‘known actuarial or statistical data’. The insurer would still be required to demonstrate that relevant statistical or actuarial data cannot be obtained, as application of the partial exemption for insurance requires the data limb and no-data limb to be applied in a strict sequence.¹¹⁹ The sequential nature of the limbs implies that for the model to be considered under the no-data limb, first it must be established no relevant data was available at the moment of contract underwriting.¹²⁰ If it is available, or can reasonably be obtained, it must not be ignored. This means that even under the no-data limb, the insurer would still need to be able to show what correlation the model uses: for example, a link between a person’s driving style and the risk they present for the purpose of car insurance.¹²¹

The second question we raise, as to the reasonableness of the discrimination, applies to both the data limb and the no-data limb, which means potential classification of the model under those limbs is less relevant. The discrimination must be reasonable having regard to the matter of data (the data limb) or other relevant factors (both limbs). In *QBE Travel Insurance v Bassanelli*, the Federal Court indicated that ‘[a]ny matter which is rationally capable of bearing upon whether the discrimination is reasonable would fall within the umbrella of relevance’.¹²² In the *Bassanelli* case, other insurers’ practices were considered relevant.¹²³ Guidelines issued by the Australian Human Rights Commission (‘*AHRC Guidelines*’) provide additional examples of ‘other relevant factors’:

- medical or other professional opinion;¹²⁴
- *relevant* information about circumstances of the particular individual seeking insurance;¹²⁵
- actuarial advice;¹²⁶ or
- insurer’s commercial judgement.¹²⁷

The *AHRC Guidelines* indicate discrimination cannot be reasonable if based on untested assumptions. Case law states the test of reasonableness is

an objective one, which requires the court to weigh the nature and extent of the discriminatory effect, on the one hand, against reasons advanced in favour

¹¹⁹ *AHRC Guidelines* (n 115) 6.

¹²⁰ ADWG (n 101) 17.

¹²¹ Cf the study described in Gert Meyers and Ine Van Hoyweghen, “‘Happy Failures’: Experimentation with Behaviour-Based Personalisation in Car Insurance” (2020) 7(1) *Big Data and Society* 1.

¹²² *Bassanelli* (n 115) 99 [53].

¹²³ *Ibid* 97 [41].

¹²⁴ *AHRC Guidelines* (n 115) 11–12 [4.2(a)], 13 [4.2(c)].

¹²⁵ *Ibid* 12–13 [4.2(b)].

¹²⁶ *Ibid* 13 [4.2(d)].

¹²⁷ *Ibid* 14 [4.2(f)].

of the requirement or condition on the other. All circumstances of the case must be taken into account.¹²⁸

It follows that potentially discriminatory decisions may require careful consideration and must take into account:

- ‘practical and business implications’;
- ‘whether less discriminatory options were available’;
- ‘the individual’s particular circumstances’;
- legislative objects, especially the object of eliminating discrimination as far as possible; and
- ‘all other relevant factors of the particular case’.¹²⁹

Furthermore, for disability discrimination, an insurer who imputes a disability merely from a medical consultation is not acting reasonably under the *AHRC Guidelines*.¹³⁰

In the case of indirect discrimination, the question of reasonableness is particularly complex. As discussed above, if a protected attribute is a predictive variable, even if it is not included in the data, the model will approximate it from other available data.¹³¹ Lack of information on the protected attribute in the dataset will only make it more complicated to de-bias the model. Due to the opacity and complexity of AI models, it may be particularly challenging to establish whether a model discriminates at all, and if it does, whether the discrimination is unreasonable and therefore unlawful.

Let us consider a hypothetical example. Statistically speaking, an insurer’s model proposes higher car insurance prices for people with mental health issues, such as depression. Is this unreasonable discrimination? People with depression *may* be higher risk, but the feature the model is considering when setting the price is the accident history, and not a person’s mental health. Therefore, it may be shown that insureds with depression, with better accident histories, receive the same pricing outcomes as insureds without mental health issues, and healthy insureds with bad accident histories pay higher prices, as do insureds with depression with similar accident histories. It may statistically be the case that people with depression more often have worse accident histories. However, accident history is a much better predictor of risk than mental health, so the model’s decision is not based on mental health. Maybe, if accident history was unavailable, the model would take into account mental health as a good predictor of accident history, but availability of more granular data prevents it. However, now imagine that depression is truly predictive of a higher risk, irrespective of a person’s accident history. In such case, a ML model would actually be discriminating against people with a disability. An insurer would then need to show this discriminatory outcome is reasonable; that is,

¹²⁸ *Bassanelli* (n 115) 99 [51], quoting *Secretary, Department of Foreign Affairs v Styles* (1989) 88 ALR 621, 623.

¹²⁹ *AHRC Guidelines* (n 115) 7, in the context of disability discrimination, but it can be extrapolated to all types of discrimination.

¹³⁰ *Ibid* 17.

¹³¹ See Lattimore et al (n 73) 30–34 (Scenario 1).

that an insured's mental health is not only a good predictor of car insurance claims, but it is a truly predictive variable in itself.¹³²

One problem with indirect discrimination is difficulty of observation. Anti-discrimination laws require the consumer actively to request explanation as to why they are denied insurance, or why the policy (premium or cover) is on less advantageous terms. The right is useless unless the consumer is aware of the discrimination. However, indirect discrimination may be very difficult to observe (as discussed in the context of the Apple example above), a problem that is exacerbated in the case of opaque ML models. This might mean that using such models could contravene anti-discrimination laws, as it would make it considerably more difficult for consumers to question the reasonableness of discrimination.

Insurers' historical lack of compliance makes the issue even more problematic. Cases of insurance discrimination against people with mental health conditions illustrate this well. The recent inquiry into discrimination in the travel insurance industry by the Victorian Equal Opportunity and Human Rights Commission has shown that such discrimination is a systemic issue.¹³³ The Commission investigated three major insurers, making up almost 40% of Australia's travel insurance industry, over eight months.¹³⁴ It found over 365,000 policies were issued unlawfully discriminating against people with mental health conditions.¹³⁵ Despite *Ingram*,¹³⁶ which held blanket mental health insurance exclusions constituted unlawful discrimination, such practices are still widespread. Industry attitudes towards mental health disorders seem to be changing, as indicated for instance by adoption of the new General Insurance Code of Practice.¹³⁷ Nonetheless, cases of deliberate insurance discrimination may still be prevalent and potentially exacerbated by algorithmic decision-making.

Algorithmic bias in combination with opacity, increases the risk of unlawful discrimination in an insurance context. This is especially concerning since various inquiries, including the Royal Commission, have demonstrated many insurers do not satisfactorily comply with existing rules.¹³⁸ It needs to be carefully considered by regulators and lawmakers, in particular regarding the possible introduction of clear rules requiring explainability of a model's decision, and of restrictions on the use of unexplainable models. A useful example of how such rules could operate comes from overseas. NYSDFS issued a Circular Letter requiring life insurers using

¹³² ADWG (n 101) 29–30.

¹³³ See Victorian Equal Opportunity and Human Rights Commission, *Fair-Minded Cover: Investigation into Mental Health Discrimination in Travel Insurance* (Report, 12 June 2019).

¹³⁴ *Ibid* 11.

¹³⁵ *Ibid*.

¹³⁶ *Ingram* (n 117).

¹³⁷ Insurance Council of Australia, General Insurance Code of Practice (updated 5 October 2021) pt 9, especially s 104 <<https://insurancecouncil.com.au/code-of-practice/>>.

¹³⁸ See eg *Royal Commission Report* (n 8) vol 2, 298–301, 313–15, 327–8, 344–7, 378–9, 385–90, 408–13, 425–30, 441–3, 453–4.

external sources of data to explain the rationale of their underwriting decision.¹³⁹ They cannot 'rely on the proprietary nature of a third-party vendor's algorithmic processes to justify the lack of specificity related to an adverse underwriting action'.¹⁴⁰ Such an approach should limit insurers' use of data to only such features with a causal influence on risk, and would thus require provision of an explanation if discriminatory treatment is suspected.

IV Insured's Duty of Disclosure

A Information Asymmetry

Before the *ICA*, general insurance contracts in Australia followed common law principles focusing on the insured's pre-contractual duty of disclosure.¹⁴¹ The insured's duty of disclosure, or at least the duty not to make misrepresentations to the insurer, is characteristic of insurance contracts in most jurisdictions.¹⁴² This is because, as Lord Mansfield put it in the landmark case of *Carter v Boehm*:

Insurance is a contract upon speculation. The special facts, upon which the contingent chance is to be computed, lie most commonly in the knowledge of the insured only: the under-writer trusts to his representation, and proceeds upon confidence that he does not keep back any circumstance in his knowledge, to mislead the under-writer into a belief that the circumstance does not exist, and to induce him to estimate the risque, as if it did not exist.¹⁴³

Information asymmetry between an insured and an insurer has always favoured the insured, who has, at least theoretically, all information needed to calculate risk. It has therefore been more efficient to require the insured to disclose to the insurer all facts relevant to their risk, rather than expect the insurer to investigate each potential insured. The rationale for the insured's duty of disclosure goes further than allowing insurers to price risk correctly,¹⁴⁴ but extends to preventing fraud and exploitation of information imbalance by insureds.¹⁴⁵

The paradigm, however, may be changing due to an ever-increasing creation and availability of digitalised data about consumers. Privacy protection aside, insurers are now able to collect consumers' data from external, or 'non-traditional'

¹³⁹ NYSDFS, 'RE: Use of External Consumer Data and Information Sources in Underwriting for Life Insurance', *Insurance Circular Letter No. 1 (2019) to All Insurers Authorized to Write Life Insurance in New York State* (Circular Letter, 18 January 2019) ('NYSDFS Circular Letter').

¹⁴⁰ *Ibid* II.B.

¹⁴¹ In addition to state legislation, such as the *Insurance Act 1902* (NSW), and Imperial statutes such as the *Life Assurance Act 1774* (Imp), *Fires Prevention (Metropolis) Act 1774* (Imp) and *Marine Insurance Act 1788* (Imp) all repealed by *ICA*.

¹⁴² Anthony A Tarr and Julie-Anne Tarr, 'The Insured's Non-disclosure in the Formation of Insurance Contracts: A Comparative Perspective' (2001) 50(3) *International and Comparative Law Quarterly* 577, 577.

¹⁴³ *Carter v Boehm* (1766) 3 Burr 1905; 97 ER 1162, 1164 [1909].

¹⁴⁴ Robin Bowley, 'Transparency in the Insurance Contract Law of Australia' in Pierpaolo Marano and Kyriaki Noussia (eds), *Transparency in Insurance Contract Law* (Springer International Publishing, 2019) 549, 552.

¹⁴⁵ McGurk (n 3) 141.

sources; that is, sources different from the proposal forms that consumers typically complete for the purposes of underwriting contracts. These may include all sorts of smartphone applications, consumers' social media presence, website cookies, smart homes,¹⁴⁶ healthcare and fitness devices,¹⁴⁷ cars,¹⁴⁸ and public surveillance devices with facial recognition capabilities, to name just a few.

AI tools make it possible and commercially feasible to analyse large amounts of personal data in order to extract meaningful features and ultimately evaluate insureds' risk; although we cannot confirm this is already occurring in practice at a large scale. However, there is evidence that US insurers are already accessing potential insureds' social media accounts for the purpose of underwriting of life insurance contracts.¹⁴⁹ Traditionally understood information asymmetry between the parties to an insurance contract is affected, as an insurer may obtain relevant information about the insured without asking them to provide it. Therefore, some argue the insured's duty of disclosure should be significantly restricted or even reversed.¹⁵⁰

Interestingly, a similar concern regarding information asymmetry and power imbalances between parties in consumer insurance contracts was shared by the Royal Commission. This concern resulted in Recommendation 4.5, 'Duty to take reasonable care not to make a misrepresentation to an insurer':

Part IV of the *Insurance Contracts Act* should be amended, for consumer insurance contracts, to replace the duty of disclosure with a duty to take reasonable care not to make a misrepresentation to an insurer (and to make any necessary consequential amendments to the remedial provisions contained in Division 3).¹⁵¹

The Royal Commission's reasons for recommending an overhaul of the insured's duty of disclosure did not consider ML models. Other concerning practices of insurance businesses motivated the Royal Commission's proposed reform. Commissioner Hayne considered that the insured's duty of disclosure in consumer insurance contracts (set out in s 21(1) of the *ICA*) placed a disproportionate burden on the insured. Section 21(1) provides general guidance as to what the insured must disclose: every matter relevant to the insurer's decision whether to accept risk, which the insured knows is relevant, or a reasonable person in the circumstances could be expected to know to be relevant. The insured's knowledge of relevance or

¹⁴⁶ Allianz and Deutsche Telecom provide a service whereby a smart home alerts the insurer if damage occurs: OECD (n 67) 12–13.

¹⁴⁷ Fitbit has a partnership with United Health in the US, and UK insurer Vitality collects activity data via Apple watches and offers discounts and direct contributions: OECD (n 67) 13. PitPat offers a similar service for pet insurance: OECD (n 67) 13.

¹⁴⁸ In Italy, over 2 million vehicles have been fitted with devices allowing for tracking of speed, braking, acceleration, and times of day when the vehicle is used: OECD (n 67) 12.

¹⁴⁹ NYSDFS Circular Letter (n 139).

¹⁵⁰ Julie-Anne Tarr, 'Data Profiling and Insurance Law' by Brendan McGurk' (2019) 30(2) *Insurance Law Journal* 131, 131.

¹⁵¹ *Royal Commission Report* (n 8) vol 1, 32.

materiality of the matters has been extensively discussed by scholars and judges.¹⁵² It is predominantly an objective test, focusing on what a reasonable insured would understand as relevant. It means an insured, even if asked specific questions by an insurer, must also *volunteer* other information they know to be material to the risk.

However, the Royal Commission inquiry indicated such an approach might be outdated for consumer insurance. The TAL Life Limited ('TAL') case study cited in the Final Report of the Royal Commission¹⁵³ demonstrated, in Commissioner Hayne's words, 'the breadth and depth of the gap between what a consumer knows and what an insurer knows'.¹⁵⁴ TAL's handling of claims made by three insureds under income protection policies was examined by the Royal Commission. The second insured's case provides an interesting illustration of how information collection by insurers may constitute an attempt to refuse cover, rather than correct evaluation of risk. The insured was diagnosed with cancer shortly after taking out a TAL income protection policy. The insurer, trying to find a reason for contract avoidance, sought a retrospective underwriting opinion in relation to some symptoms, which may have been indicative of the cancer and were experienced by the insured prior to entering the policy. The potential argument in the insurer's favour was that had those symptoms been disclosed by the insured, the policy would have been refused.¹⁵⁵ It was admitted that TAL would review insureds' claims and pre-contractual information provided in a form of 'fishing expedition', collecting all information available, including irrelevant material.¹⁵⁶ Clearly, use of sophisticated technology, and collecting consumers' data from various sources would make such conduct much easier for insurers. Availability of consumers' data, for example their daily shopping (as collected through retailers' loyalty programs) or their Internet browser searches and other data collected through cookies,¹⁵⁷ could provide information to insurers potentially letting them avoid claims in similar circumstances.

This potential use of technology was not discussed by the Royal Commission,¹⁵⁸ as the investigated practices were historical and use of the technology in question did not yet appear to be widespread. Nevertheless, the accessibility of consumers' personal data paired with algorithmic decision-making could exacerbate issues identified by the Royal Commission. Therefore, the Royal Commission's Recommendation 4.5, and law reform proposals put forward in response and finally adopted (discussed in Part IV(B)) need to be evaluated from the point of view of AI and Big Data use by insurers.

¹⁵² See, eg, Rob Merkin, 'What Does an Assured 'Know' for the Purpose of Pre-Contractual Disclosure?' (2016) 27 *Insurance Law Journal* 157.

¹⁵³ *Royal Commission Report* (n 8) vol 2, ch 4.

¹⁵⁴ *Ibid* vol 1, 297.

¹⁵⁵ *Ibid* vol 2, 341.

¹⁵⁶ *Ibid* vol 2, 339.

¹⁵⁷ Brigid Richmond, *A Day in the Life of Data: Removing the Opacity Surrounding the Data Collection, Sharing and Use Environment in Australia* (Consumer Policy Research Centre Report, May 2019) 15, 30, 34–5.

¹⁵⁸ Cf McGurk (n 3) 123.

The Royal Commission concluded that the duty to take reasonable care not to make a misrepresentation to an insurer is more appropriate, and less complex, than the current duty applicable to consumer contracts. Such a duty for the consumer places the onus on the insurer to ask the right questions to obtain relevant information as to risk.¹⁵⁹ The Australian Government agreed with the Royal Commission, proposing to amend the duty of disclosure for consumers in order to protect them from claims refusal due to inadvertent omissions or insurers' failure to ask appropriate questions.¹⁶⁰ This new approach has been implemented into legislation, and the insured's duty of disclosure in consumer insurance contracts has been replaced by the insured's duty to take reasonable care not to make a misrepresentation to the insurer.¹⁶¹ The changes apply to all consumer insurance contracts entered into, renewed or varied on or after 5 October 2021.¹⁶²

B *Insured's Misrepresentation to the Insurer*

What does the overhaul of the insured's duty of disclosure mean? In consumer insurance contracts (when the insurance is obtained for personal, domestic or household purposes),¹⁶³ an insured will now have a duty to take reasonable care not to make a misrepresentation to the insurer before the contract is entered into.¹⁶⁴ This includes also life insurance contracts, as well as contracts formerly belonging to the 'eligible contracts' category.¹⁶⁵

Insurers are protected against insureds' misrepresentation through remedies of contract avoidance or liability reduction to the amount placing the insurer in a position they would have been in had the misrepresentation not been made.¹⁶⁶ Section 28(1) of the *ICA* provides that remedies are not available to the insurer if, even though a misrepresentation occurred, it would have entered the contract on the same terms. The onus of proof is on the insurer, who must demonstrate they would not have entered the contract had they known what the insured misrepresented. Case law shows clear guidelines used by insurers as to whether they would accept various risks can be helpful in proving what they would or would not have done.¹⁶⁷ Use of (explainable) statistical models (ML or otherwise) could also be of assistance

¹⁵⁹ *Royal Commission Report* (n 8) vol 1, 300.

¹⁶⁰ Draft Explanatory Memorandum, Financial Sector Reform, Hayne Royal Commission Response – Protecting Consumers (2020 Measures) Bill 2020 (Cth), rec 4.5 (Duty of Disclosure to Insurer) [1.10]–[1.11].

¹⁶¹ *Financial Sector Reform (Hayne Royal Commission Response) Act 2020* (Cth) sch 2 pt 2, amending the *ICA* with effect from 1 January 2021.

¹⁶² *Ibid* sch 2 s 37.

¹⁶³ *Insurance Contracts Act 1984* (Cth) s 11AB(1) ('ICA'). See also sub-ss (2)–(3).

¹⁶⁴ *Ibid* s 20B(1).

¹⁶⁵ *Financial Sector Reform (Hayne Royal Commission Response) Act 2020* (Cth) sch 2 s 9 repealed *ICA* (n 163) s 21(A).

¹⁶⁶ *ICA* (n 163) ss 28(2)–(3).

¹⁶⁷ Julie-Anne Tarr, 'Insurance Contract Disclosure: An Uncertain Balance' (2015) 26(2) *Insurance Law Journal* 109, 115–16. See, eg, *Michail v Australian Alliance Insurance Co Ltd* (2014) 67 MVR 21, especially 29 [31].

regarding such proof, as it would be possible to provide the new, previously undisclosed information item to the model to check how the outcome changes.

In general terms, we consider that eliminating the consumer's duty of disclosure is a positive development in the age of AI and Big Data. The new rules state that whether an insured has taken reasonable care not to make a misrepresentation is to be determined considering all the relevant circumstances.¹⁶⁸ Although it should be assumed, without more, that the insured is an average person with no special skills or knowledge, the test is ultimately subjective. If the insurer knew, or ought to have known, about particular characteristics or circumstances of the insured individual, these characteristics or circumstances must be considered in determining whether the insured has taken reasonable care not to make a misrepresentation to the insurer.¹⁶⁹ In this context, an insurer's wide collection of data may have important consequences. It may lower the standard of care required of the consumer to discharge their duty, such as in cases where data collected by the insurer indicates that they have a certain type of vulnerability or disability. Therefore, the situations in which an insurer is deemed to have knowledge of the insured's circumstances will need to be clarified.

The question is also how the standard of care required of the insured for the purpose of making relevant representations to the insurer would be affected by use of AI and Big Data tools by an insurer. Theoretically, this could be relevant for the purpose of s 20B(2) of the *ICA*. The promises of profit brought about by the technology use are likely to incentivise data collection by insurers, including collecting detailed data about a consumer,¹⁷⁰ as well as attempts to make sophisticated inferences.¹⁷¹ Use of sophisticated ML models and Big Data collection drives the information asymmetry between the parties even further in an insurer's favour. Therefore, the overhaul of an insured's duty of disclosure in consumer insurance contracts is a welcome development, aiming at reflecting and remedying the imbalance of power and information asymmetry between the parties. Consequently, the standard of care required from the insured should also reflect this. The rules are new, and untested in this context, so we can only offer a general interpretation. Insurers should not be able to avoid paying a claim on the basis of an insured's misrepresentation as to the risk, if the insurer knew the insured's risk circumstances from collected data and subsequent analytics. This should be, at least partly, captured by s 28 *ICA*, with remedies unavailable to the insurer if they would have entered the contract on the same terms anyway.

The question, then, is when an insurer would be understood as 'knowing' something. In terms similar to s 21(2) of the *ICA*, which now does not apply to consumer contracts, it could be construed that an insurer knows matters 'known to an appropriate officer or agent of the insurer or contained in current official

¹⁶⁸ See *ICA* (n 163) s 20B(2)–(3) for examples of such relevant circumstances.

¹⁶⁹ *Ibid* s 20B(4).

¹⁷⁰ Marshall Allen, 'Health Insurers Are Vacuuming up Details about You: And It Could Raise Your Rates', *NPR* (online, 17 July 2018) <<https://www.npr.org/sections/health-shots/2018/07/17/629441555/health-insurers-are-vacuuming-up-details-about-you-and-it-could-raise-your-rates>>.

¹⁷¹ See the explanation of inferences in ML models in Part II(A)(4) above.

records'.¹⁷² Courts have held that insurers' knowledge could not automatically be inferred based solely on insurers' access to some written information (for example, a newspaper extract) in paper-based general files held by the insurer.¹⁷³ Despite this, in our view the use of digital information systems by modern insurers will mean the benefit of ignorance should not be easily available.¹⁷⁴ This argument is even stronger when AI and Big Data tools are used, especially inferentially.

V Insurer's Obligations

A Information Duties

In the light of all the issues we have discussed relating to the insured's duty of disclosure, we must now consider an insurer's obligations towards the insured. These obligations originate in the overarching duty of utmost good faith out of s 13 of the *ICA*, as well as in other specific statutory provisions.

As set out in Part II(A), the opacity of ML models is potentially problematic, especially considering the significant limits on local explainability. Opacity may cause significant harm to consumers, as algorithmic bias or discriminatory outcomes may be effectively unobservable by affected individuals. Also, in the context of disclosure duties, insureds may not know, justifiably so, what matters would be relevant to algorithmic risk assessment. If refused cover, or offered a different cover or premium than expected, a lack of a local explanation will deprive the consumer of (potential) useful feedback from the insurer: feedback that could help them change their practices to obtain a better deal. The inability to trace this feedback to particular consumer features renders information on premiums and cover of doubtful value given unidentified factors leading to the result may change at any time. The consumer's right to know which features affected the underwriting decision and in what way should be mirrored by a duty on the insurer to provide them with such information.

There are various concrete duties imposed on insurers requiring certain conduct or information provision to insureds, but only some of those duties may play a role in the context of use of Big Data and AI analytics.¹⁷⁵ Of particular interest are s 75 of the *ICA* and ss 160–3 of the General Insurance Code of Practice. Section 75 of the *ICA* provides an insured with a right to request written reasons from an insurer who:

- rejects a potential insured;
- cancels a contract;
- does not offer renewal; or
- offers insurance cover on terms less advantageous than terms they would otherwise offer.¹⁷⁶

¹⁷² Julie-Anne Tarr, "Knowledge" and Pre-Contract Disclosure under the *Insurance Contracts Act* (2018) 46(6) *Australian Business Law Review* 355, 367.

¹⁷³ Ibid, see *Commercial Union Assurance Co of Australia Ltd v Beard* (1999) 47 NSWLR 735, 735. For the UK, see, eg, *Malhi v Abbey Life Insurance* [1996] LRLR 237, 237.

¹⁷⁴ Tarr, "Knowledge" and Pre-contract Disclosure under the *Insurance Contracts Act* (n 172) 367.

¹⁷⁵ As discussed in Part I, data protection laws are outside the scope of this article.

¹⁷⁶ Cf *Sex Discrimination Act* (n 110) s 41(1)(e).

Sections 160–63 of the Code of Practice also deal with the insured's access to information. Insurers are under a duty to provide an insured with any information relied on in assessing their application for insurance cover, in handling the claim, or in responding to their complaint. An insurer may refuse to provide access to the requested information, but they cannot do so unreasonably.

As discussed above in the context of anti-discrimination laws, the consumer must proactively request information in writing, which constitutes a significant barrier. Even if they do so, the provision does not specify the nature of data to which the individual would be entitled.¹⁷⁷ It was argued in the context of use of genetic information for life insurance underwriting that the consumer should be entitled to 'an explanation, in layman's terms, of the reasons for the unfavourable underwriting judgment and the actuarial basis for that decision'.¹⁷⁸ However, both s 75 of the *ICA* and the General Insurance Code of Practice do not require this. Furthermore, disclosing data influencing underwriting decisions is costly and time-consuming even when ML decision-making is not involved,¹⁷⁹ and it also risks compromising commercial confidentiality.¹⁸⁰ While disclosing *exact* data underpinning an underwriting decision to a consumer may not be necessary, there is no relevant guidance on the level of detail required by either s 75 of the *ICA* or the Code of Practice.

B Other Obligations Regarding Insurers' Conduct towards Consumers

Various provisions require insurers to comply with a specified high standard of conduct, especially when they are dealing with consumers. Apart from those already mentioned, there is a series of rules arising from insurance contracts being financial products, set out in the *Corporations Act 2001* (Cth) ('*Corporations Act*') and *Australian Securities and Investments Commission Act 2001* (Cth) ('*ASIC Act*').¹⁸¹ These require insurers to:

- provide their services efficiently, honestly and fairly;¹⁸²

¹⁷⁷ Australian Law Reform Commission ('ALRC'), *Essentially Yours: The Protection of Human Genetic Information in Australia* (Report No 96, May 2003) ('*Human Genetic Information Report*') vol 2, 717–18 [27.71]–[27.74] citing, eg, Margaret Otlowski, Submission No G159 to the ALRC, *Joint Inquiry into the Protection of Human Genetic Information* (24 April 2002) ('Otlowski, Submission') and Institute of Actuaries of Australia, Submission No G224 to the ALRC, *Joint Inquiry into the Protection of Human Genetic Information* (29 November 2002).

¹⁷⁸ *Human Genetic Information Report* (n 177) vol 2, 719 [27.80] citing Centre for Law and Genetics, Submission No G048 to the ALRC, *Joint Inquiry into the Protection of Human Genetic Information* (14 January 2002).

¹⁷⁹ *Human Genetic Information Report* (n 177) vol 2, 720 [27.81] citing Investment and Financial Services Association, Submission No G244 to the ALRC, *Joint Inquiry into the Protection of Human Genetic Information* (19 December 2002).

¹⁸⁰ *Human Genetic Information Report* (n 177) vol 2, 720 [27.81], citing Otlowski Submission (n 177).

¹⁸¹ *Australian Securities and Investments Commission Act 2001* (Cth) s 12BAA(7)(d) ('*ASIC Act*'); *Corporations Act 2001* (Cth) ss 764A(1)(d)–(e) ('*Corporations Act*').

¹⁸² *Corporations Act* (n 181) s 912A(1)(a). For a detailed discussion of the concept, see Peter Mann and Stanley Drummond, 'Utmost Good Faith, Unconscionable Conduct and Other Notions of Fairness: Where Are We Now?' (2017) 29(1) *Insurance Law Journal* 1, 47–52.

- comply with other obligations of Australian financial services licensees;¹⁸³
- prohibit misleading or deceptive conduct;¹⁸⁴ and
- prohibit unconscionable conduct.¹⁸⁵

Standards imposed by these rules, broadly speaking, require fair and honest conduct by financial services providers, similar to an insurer's duty of utmost good faith.¹⁸⁶ Although these standards are relevant in the context of use of AI and Big Data tools, this article's main focus is on the duty of utmost good faith, 'a foundation stone and guiding principle of insurance and insurance law'.¹⁸⁷

C *Insurer's Duty of Utmost Good Faith*

An insurance contract is a special type of contract in the terms of protections offered to both contractual parties. Although historically the focus was on insureds' duties of disclosure, insurers also have obligations stemming from utmost good faith towards insureds. *Carter v Boehm* held that '[g]ood faith forbids either party by concealing what he privately knows, to draw the other into a bargain, from his ignorance of that fact, and his believing the contrary.'¹⁸⁸ This also includes the insurer, as 'the policy would equally be void, against the under-writer, if he concealed; as, if he insured a ship on her voyage, which he privately knew to be arrived: and an action would lie to recover the premium.'¹⁸⁹

Underwriting of insurance contracts is highly regulated by statute, and specific provisions do offer some (limited) protection for consumers against harms resulting from use of AI models for analysing consumers' data. Section 13 of the *ICA* restates the common law principle in *Carter v Boehm*, implying a term in insurance contracts requiring all parties to act, in respect of any matter arising under or in relation to the contract, with the utmost good faith. The duty has been divided into four 'quadrants', and covers both pre-contractual and post-contractual phases.¹⁹⁰ In this section, we focus on pre-contractual operation of the duty in respect of the insurer's use of AI and Big Data technologies, including how pre-contractual use of AI and Big Data may relate to later decisions regarding payment of claims.

The insurer's duty of utmost good faith at the pre-contractual stage applies to all aspects of the parties' relationship. The scope of the duty has been judicially described as: 'an insurer's statutory obligation to act with utmost good faith may

¹⁸³ *Corporations Act* (n 181) s 912A(1). For a detailed discussion, see Mann and Drummond (n 182) 52–4.

¹⁸⁴ *ASIC Act* (n 181) s 12DA(1); *Corporations Act* (n 181) s 1041H(1).

¹⁸⁵ *ASIC Act* (n 181) ss 12CA–12CB; *Corporations Act* (n 181) s 991A(1). See also Mann and Drummond (n 182) 20–47.

¹⁸⁶ Mann and Drummond (n 182) 1–2.

¹⁸⁷ *Ibid* 2.

¹⁸⁸ *Carter v Boehm* (n 143) 1164 [1910].

¹⁸⁹ *Ibid* 1164 [1909].

¹⁹⁰ Peter Mann, 'The Elusive Second Quadrant of Utmost Good Faith: What Is the Scope of an Insurer's Pre-Contractual Duty of Utmost Good Faith?' (2016) 27 *Insurance Law Journal* 176, 176; Bowley (n 144) 549; Mann and Drummond (n 182) 8–9.

require an insurer to act, consistently with commercial standards of decency and fairness, with due regard to the interests of the insured'.¹⁹¹ The insurer's obligation to act following the standards of decency, fairness and honesty¹⁹² in the pre-contractual phase can be divided into two components: the duty of disclosure towards the insured, and the insurer's conduct beyond the disclosure. Both are relevant to the use of AI models in consumer insurance contracts.

First, will insurers need to inform the prospective insured what data is used for underwriting of the contract, and how data is processed? In the context of English law, it has been argued that insurers who use predictive models should be required to disclose all matters affecting risk evaluation, so that insureds could understand the basis upon which proposed cover has been offered.¹⁹³ Australian law, however, although stemming from the same common law principles, has evolved differently, and the statutory duty of s 13 of the *ICA* is different to the common law. English law therefore offers limited assistance.

Case law relevant to the insurer's pre-contractual duty of utmost good faith demonstrates it is not absolute. An insurer's knowledge of an insured's underinsurance will not necessarily amount to breach of the duty, as set out in *Kelly v New Zealand Insurance Co Ltd*.¹⁹⁴ In *Kelly*, the insurer knew the insured's house was furnished with antiques and other expensive items, yet they accepted an increase in premium without explaining to the insured the consequences of failing to provide a list of items. However, the Court considered that only the insured knew what specific items were in the residence and their overall value. Consideration of the insurer's knowledge was important, as they had a loss adjuster's report referring to several expensive items in the insured's residence. The Court's decision that the insurer was not in breach of the duty of utmost good faith seems to indicate this report was seen as insufficiently specific. And the insured refused to provide a list of items, possibly due to concerns about tax authorities finding out about his house's contents.¹⁹⁵

Could the findings in *Kelly* be extrapolated to issues of collection and use of digital consumer data? The problem lies in the fact that AI models for extracting and inferring relevant data operate on probabilities and, in the current state of technological development, cannot, in many cases, provide specific concrete 'knowledge' regarding the circumstances of insureds. The question is, how detailed and specific would an insurer's (undisclosed) 'knowledge' need to be at the time of the policy being granted, for it to be considered in breach of the duty of utmost good faith when it subsequently uses this knowledge to deny an insured's claim?

In the age of Big Data and AI, individual data footprints have increased exponentially due to growth of social media and connected consumer devices, affordable technology to collect and process data is readily available, and data

¹⁹¹ *CGU Insurance Limited v AMP Financial Planning Pty Ltd* (2007) 235 CLR 1, 12 [15].

¹⁹² *Ibid* 42–3 [130]; *Australian Securities and Investments Commission v Youi Pty Ltd* [2020] FCA 1701, [7] ('*ASIC v Youi*').

¹⁹³ *McGurk* (n 3) 164.

¹⁹⁴ *Kelly v New Zealand Insurance Co Ltd* (1996) 130 FLR 97, 111–12 ('*Kelly*').

¹⁹⁵ *Ibid* 97–8.

brokerage still constitutes a multi-billion dollar industry despite recent setbacks such as data breaches and restriction of access by platforms.¹⁹⁶ These conditions have significant potential to affect information and power asymmetry between insurers and insureds, as insurers have unprecedented access and means to analyse insureds' digital data.¹⁹⁷ The insurers' duty of utmost good faith requires them to act in a fair, reasonable, decent way, with regard to the interests of insureds.¹⁹⁸ This does not require insurers to put insureds' interests before their own.

However, we believe that the duty of utmost good faith, in the context of unprecedented technical advantages, would require insurers at least to disclose to insureds what was discovered through data analysis. For example, in a factual scenario similar to *Kelly*, an insurer's duty to act in utmost good faith would imply the need to warn the insured that valuable items in their home, about which the insurer knows, would not be covered if a detailed list is not provided. The use of technological advancements provides great advantage to insurers. If they can rely on data accessed for underwriting purposes, they should be held accountable when it comes to paying claims. We argue that to act with utmost good faith means a quid pro quo ought to apply: if insurers use Big Data analytics to price the risk, this will affect their duty to pay claims.¹⁹⁹ The main problem is the opacity around the use of ML tools.²⁰⁰ Therefore, in the context of the use of advanced data analytics by insurers, we consider that the minimum obligation of insurers should be disclosure of what is known and how it may affect potential claims. While this goes against the decision in *Kelly*, we argue that the use of new technologies is a significant change, and higher standards should apply.

Similar considerations apply to the second component of insurers' decent, fair and honest obligation: that is, insurers' conduct beyond disclosure. To act with *utmost* good faith means more than just to act in good faith, encompassing 'notions of fairness, reasonableness and community standards of decency and fair dealing'.²⁰¹ The courts note '[w]hile dishonest conduct will constitute a breach of the duty of

¹⁹⁶ See David Lazarus, 'Shadowy Data Brokers Make the Most of Their Invisibility Cloak', *Los Angeles Times* (online, 5 November 2019) <<https://www.latimes.com/business/story/2019-11-05/column-data-brokers>>; Steven Mendelez and Alex Pasternack 'Here Are the Data Brokers Quietly Buying and Selling Your Personal Information' *Fast Company* (online, 3 February 2019) <<https://www.fastcompany.com/90310803/here-are-the-data-brokers-quietly-buying-and-selling-your-personal-information>>; Zack Whittaker, 'Data Brokers Track Everywhere You Go, But Their Days May Be Numbered' *TechCrunch* (online, 9 July 2020) <<https://au.news.yahoo.com/data-brokers-track-everywhere-days-130038354.html>>.

¹⁹⁷ The *Privacy Act 1988* (Cth) places some restrictions on data collection, use and disclosure. It is beyond the scope of this article to provide a detailed analysis of this legislation in this context. However, there is ample evidence supporting a view that the legislation has, in practice, provided few real limits on use of consumer data: see, eg, Australian Competition and Consumer Commission, *Digital Platforms Inquiry Final Report* (Report, June 2019); Kayleen Manwaring, Katharine Kemp and Rob Nicholls, (mis)Informed Consent in Australia (Report for iappANZ, 31 March 2021) <http://handle.unsw.edu.au/1959.4/unsworks_75600>.

¹⁹⁸ *Speno Rail Maintenance Australia Pty Ltd v Metals & Minerals Insurance Pte Ltd* (2009) 226 FLR 306, 330 [144], 331 [153] (Beech AJA).

¹⁹⁹ This idea was suggested by an anonymous reviewer of this article.

²⁰⁰ See Part II(A)(3) above.

²⁰¹ *AMP Financial Planning Pty Ltd v CGU Insurance Ltd* (2005) 146 FCR 447, 475 [89].

utmost good faith, so will capricious or unreasonable conduct.²⁰² Using opaque ML models, especially when there is a demonstrable detrimental effect on insureds, or prospective insureds (for example, if cover is denied), could be considered capricious and unreasonable. However, as discussed above in the context of indirect discrimination, detrimental and discriminatory effects on insureds may be unobservable for affected parties, as well as for insurers. For insurers to comply with the utmost good faith standard, it may therefore be necessary carefully to consider operation of their ML models and Big Data collection used in underwriting of consumer contracts. We argue that insurers knowingly accepting that their underwriting procedures are using opaque, unexplainable models, without efforts to control bias and procedural fairness, could be considered failing to meet the utmost good faith standard.

The discussion about usefulness of the duty of utmost good faith cannot be separated from remedies potentially available to aggrieved consumers. The common law remedy mentioned in *Carter v Boehm*, that of voiding the policy and allowing premium recovery by an insured,²⁰³ is of little use to insureds due to the nature of insurance contracts. However, breach of the statutory duty under s 13 of the *ICA* does give rise to damages, as a breach of an implied contract term that also applies to contract formation.²⁰⁴ Damages are awarded in contract.

This raises a problem. If conduct offending against the duty of utmost good faith is pre-contractual, how can an implied term be breached?²⁰⁵ If the contract is ultimately entered into, pre-contractual conduct breaching the duty of utmost good faith can result in a contract-based remedy.²⁰⁶ However, when the contract is not entered into, there is no implied term, and so there will be no contractual remedy based on breach of the duty of utmost good faith. Further obstacles to access to remedies for breach of the insurers' duty of utmost good faith include its uncertain definition, and difficulties in proving breach. Self-represented consumers before the Australian Financial Complaints Authority are unlikely to grasp fully the nature and extent of the duty and would likely fail in their argument.²⁰⁷

Section 13(2) of the *ICA* provides that a failure by a party to a contract of insurance to comply with the duty of utmost good faith is a breach of the *ICA*, attracting a civil penalty under s 13(2A).²⁰⁸ Sections 75A–75ZE detail enforcement of civil penalty provisions by the *Australian Securities and Investments Commission* ('ASIC'). Insurers who are breaching or likely to breach their utmost good faith duties are under an additional duty to self-report to ASIC under s 912D of the

²⁰² Ibid.

²⁰³ *Carter v Boehm* (n 143) 1164 [1909].

²⁰⁴ *Imaging Applications Pty Ltd v Vero Insurance Ltd* [2008] VSC 178, [54]–[55].

²⁰⁵ Mann (n 190) 176.

²⁰⁶ Ibid 180.

²⁰⁷ Ryan Natrass, 'Extending the Unfair Contract Terms Laws to Insurance Contracts: Is the Duty of Utmost Good Faith Fair Enough?' (2012) 23(3) *Insurance Law Journal* 299, 310.

²⁰⁸ Introduced through *Treasury Laws Amendment (Strengthening Corporate and Financial Sector Penalties) Act 2019* (Cth).

Corporations Act.²⁰⁹ For the duty to report to arise, the breach or likely breach must be significant, taking into account, for example, the number or frequency of similar previous breaches.²¹⁰ These rules require insurers to have a good understanding of the algorithmic decision-making processes they are using for underwriting contracts, and awareness of problems relating to bias or unexplainability potentially amounting to breach of their duty of utmost good faith. The self-reporting duty could become an important tool in the context of the use of emerging technologies and consequent compliance with financial services and insurance law.

ASIC enforcement powers are outlined in ss 915A–915J of the *Corporations Act*, and include powers to vary, suspend or cancel an insurer's financial services licence. ASIC may also issue banning orders under ss 920A–920F, prohibiting a person from providing a financial service permanently or for a specified period. ASIC will normally act only where it foresees a general benefit to the market and the public. Isolated breaches are unlikely to attract more serious penalties such as a permanent banning order.²¹¹

ASIC's powers under s 55A of the *ICA* are also important. If an insured has suffered, or is likely to suffer, damage, due to contract terms or the insurer's conduct breaching *ICA* requirements, ASIC may act against the insurer on behalf of the aggrieved party if it believes it is in the public interest to do so. ASIC may also act on behalf of a group of insureds. Public interest is evident in preventing consumers suffering damage owing to an insurer using AI and Big Data tools, given the scope of likely harm. ASIC's powers in this context can provide important assistance to consumers. However, this is not a perfect remedy, especially because the regulator acts on behalf of, and at the application of, the aggrieved party. Additionally, systemic and potentially unobservable breaches of utmost good faith duties and anti-discrimination rules may arise.

VI Conclusion

An increasing use of algorithmic decision-making for the purposes of underwriting consumer insurance will inevitably affect parties' relationships. Insurance contracts are special, being contracts on speculation, and both common law and legislation have imposed specific duties on both parties, aiming at balancing rights and burdens. Rules applicable to these contracts have been developed across centuries, but it is only recently that sociotechnical changes have brought about a need for more far-reaching interventions. Recent inquiries into the insurance industry have, unfortunately, demonstrated important shortcomings regarding the fairness of consumer treatment by some insurers. The most important issue is that current rules have been breached by insurers to refuse claims, exclude cover, or unjustifiably raise premiums. Against this background, technological advancements bring yet another challenge for the insurance industry. On the one hand, AI and Big Data tools promise unprecedented benefits in terms of costs reduction and efficiency to insurers, and

²⁰⁹ In more detail, see Mann and Drummond (n 182) 53–4.

²¹⁰ See *Corporations Act* (n 181) s 912D(1)(b).

²¹¹ See Explanatory Memorandum, Insurance Contracts Amendment Bill 2013 (Cth) [1.14].

potentially also insureds. On the other hand, the potential for consumer harm is significant. The proposal of considered solutions to these harms is beyond the scope of this article, but we offer some preliminary observations warranting further investigation.

Our analysis shows that changes in behaviour by insurers arising out of use of emerging technologies such as AI and Big Data are not wholly unregulated by existing law. The content of some consumer protection provisions specifically applicable to insurance contracts would, on their face, be adequate to safeguard consumers against some of the more egregious potential abuses by insurers. However, the Royal Commission has uncovered a serious and systemic lack of compliance with those provisions, and therefore additional incentives — punitive, persuasive, or both — are needed. The Royal Commission proposed a more interventionist approach by ASIC, and we support that call.

However, the existing law also contains significant uncertainty in its application to existing and potential new conduct by insurers using these emerging technologies. Better guidance is required on what good behaviour by insurers looks like in the new sociotechnical reality. Enforceable codes of conduct, to which industry, regulators and consumers contribute, should be useful. A co-regulatory process²¹² of this nature could go a long way to improving consumer trust in both the insurance industry and new technologies, benefitting all market players.²¹³

Any new regulatory provisions, whether by code of conduct or otherwise, must deal with the current lack of transparency and explainability of ML decision-making. These deficiencies affect consumer choice and adequacy of cover, content and timing of regulatory intervention, and quality of judicial decision-making. The prospect of mandatory human intervention in decision-making must also be considered.²¹⁴ We recognise that this is a complex problem without easy solutions, but delay potentially exacts a significant price. Insurers attracted by the prospect of more cost-efficient business models are likely to make substantial investment in these technologies — with consequent entrenched resistance to regulatory intervention increasing over time.²¹⁵

²¹² See, eg, Roger Clarke, 'Regulatory Alternatives for AI' (2019) 35(4) *Computer Law & Security Review* 398, 406–7.

²¹³ On trustworthy AI, see SA Hajkowicz, Sarvnaz Karimi, Tim Wark, Caron Chen, M Evans, Natalie Rens, Dave Dawson, Andrew Charlton, Toby Brennan, Corin Moffatt, Sriram Srikumar and K J Tong, *Artificial Intelligence: Solving Problems, Growing the Economy and Improving Our Quality of Life* (CSIRO Data61 Report, 2019) 56.

²¹⁴ OECD (n 67) 22–3.

²¹⁵ Kayleen Manwaring, 'Will Emerging Information Technologies Outpace Consumer Protection Law? The Case of Digital Consumer Manipulation' (2018) 26(2) *Competition and Consumer Law Journal* 141, 177–81.

Fighting the System: New Approaches to Addressing Systematic Corporate Misconduct

Samuel Walpole* and Matt Corrigan†

Abstract

Traditional criminal law evolved to address morally unacceptable conduct by individuals, before expanding into regulatory contexts. Classical models of corporate criminal responsibility sought to apply the individual-focused criminal law to corporate defendants. At the same time, contemporary corporations act increasingly in ways distinct from natural persons: through systems, patterns of behaviour, policies, procedure, and culture. This has led to increasing interest in the framing of offences that are better tailored to the way in which corporations act in reality. Building on the recommendations of the Australian Law Reform Commission in its *Corporate Criminal Responsibility Report*, this article considers a novel type of offence — one that criminalises systems of conduct or patterns of behaviour by corporations. We argue that system of conduct offences have the potential to enhance corporate criminal law's effectiveness, and to serve as an alternative to traditional approaches to corporate criminal liability in appropriate contexts.

Please cite this article as:

Samuel Walpole and Matt Corrigan, 'Fighting the System: New Approaches to Addressing Systematic Corporate Misconduct' (2021) 43(4) *Sydney Law Review* 489.



This work is licensed under a Creative Commons Attribution-NonDerivatives 4.0 International Licence (CC BY-ND 4.0).

As an open access journal, unmodified content is free to use with proper attribution.

Please email sydneylawreview@sydney.edu.au for permission and/or queries.

© 2021 Sydney Law Review and authors. ISSN: 1444-9528

* Barrister-at-Law; Adjunct Fellow, TC Beirne School of Law, The University of Queensland, Australia. Email: s.walpole@uq.edu.au; ORCID iD: <https://orcid.org/0000-0002-3360-5204>.

† General Counsel, Australian Law Reform Commission. Email: matt.corrigan@alrc.gov.au; ORCID iD: <https://orcid.org/0000-0003-0914-1849>.

The authors were both participants in the Australian Law Reform Commission ('ALRC') Inquiry into Corporate Criminal Responsibility. We thank Professor Elise Bant, Professor Jeannie Marie Paterson, Phoebe Tapley, Tess van Geelen and the anonymous reviewers for their comments on an earlier draft. The views expressed in this article reflect the personal views of the authors, and any errors are solely our own.

I Introduction

Current policy debates have highlighted growing appreciation of the distinctive nature of corporate, as opposed to individual, criminality. Corporate criminal responsibility has traditionally sought to adapt individual-focused models of criminal liability to corporate defendants by using attribution methods, based on the recognition that a corporation is an artificial legal construct that acts *through* individuals. While a corporation must act through individuals, contemporary corporate action is not limited to one, or a handful of individuals. Corporations can act — and fail to act — in ways that are different from individual natural persons. Collective decision-making, computer programs, systems of conduct, patterns of behaviour, policies, procedures, and culture can all represent acts and omissions by corporations.¹ Traditional approaches to corporate responsibility, which are based on the attribution of acts and mental states of natural persons to corporations, do not easily accommodate these types of corporate action.

In recent years, various law reform initiatives have been proposed to address corporate misconduct arising from deficient systems, practices, policies, and cultures. Such initiatives have been driven by public disquiet about particular examples of misconduct, ranging from financial institutions charging their clients for services never delivered, through the push for stronger laws relating to industrial manslaughter, to calls for the criminalisation of wage theft.² Relatedly, there is wider recognition of the need to tailor more appropriately the criminal law to contemporary corporate defendants, due to the limitations of responsibility based on traditional attribution principles. This has resulted in the enactment of an increasing number of offences tailored to the reality of corporate action. These types of offences can better reflect the characteristics of the contemporary corporate form. Examples of such offences, which eschew reliance on traditional principles of attribution, include ‘duty-based’ offences relating to workplace health and safety and to the regulation of heavy vehicles,³ and the ‘failure to prevent offence’ that has been proposed for

¹ The corporate culture provisions in pt 2.5 of the *Criminal Code Act 1995* (Cth) sch (‘*Criminal Code* (Cth)’) are a recognition of the role of corporate culture in facilitating — or preventing — criminal conduct by a corporation: *Criminal Code* (Cth) s 12.3(2)(c)–(d).

² A number of Australian jurisdictions have recently taken steps to address wage theft. In June 2020, Victoria passed the *Wage Theft Act 2020* (Vic). The Queensland Government passed reforms to the *Criminal Code Act 1899* (Qld) sch (‘*Criminal Code* (Qld)’) in September 2020. At the Commonwealth level, the Fair Work Amendment (Supporting Australia’s Jobs and Economic Recovery) Bill 2020 (Cth) proposed the introduction of an offence where an ‘employer dishonestly engages in a systematic pattern of underpaying one or more employees’, however the offence was omitted from the legislation that was ultimately passed.

³ See, eg, *Work Health and Safety Act 2011* (Cth) (‘*WHS Act* (Cth)’); *Heavy Vehicle National Law Act 2012* (Qld). See further Australian Law Reform Commission, *Corporate Criminal Responsibility* (Final Report No 136, April 2020) 321–7 [7.178]–[7.196] (‘*ALRC Corporate Criminal Responsibility Report*’).

foreign bribery.⁴ In particular contexts, offences like these have real potential to secure corporate accountability and improve corporate culture.⁵

Within the context of such law reforms, this article explores a novel type of offence for addressing systematic corporate misconduct first put forward, in a particular form, by the Australian Law Reform Commission ('ALRC') in its *Corporate Criminal Responsibility Report*.⁶ The ALRC termed such offences 'system of conduct offences', as they are based on the 'system of conduct or pattern of behaviour' concept used in the statutory unconscionability civil regulatory provisions.⁷ The ALRC recommended the enactment, in appropriate contexts, of a type of system of conduct offence to criminalise contraventions of prescribed civil penalty provisions that constitute a system of conduct or pattern of behaviour by a corporation.⁸ This article builds on the work of the ALRC, and suggests that system of conduct offences might be appropriate beyond circumstances involving multiple civil penalty contraventions, as a means of overcoming the limitations of traditional models of corporate criminal responsibility. This article proposes a more broadly applicable model of criminal offence that responds to corporate systems and patterns of behaviour that result in criminal offending and that, due to these characteristics, could provide the foundation for corporate criminal liability without the application of traditional principles of attribution.

Central to both the ALRC's recommendation and the broader offence model that we explore in this article is the concept of a system of conduct or pattern of behaviour. The ALRC adopted this concept as the foundation for its proposed system of conduct offence due to the 'developing body of jurisprudence' surrounding the meaning of this concept.⁹ Part IV(B) below discusses the existing case law as to what constitutes a 'system' or 'pattern'. Put shortly, however, 'a "system" connotes an internal method of working; a "pattern" connotes the external observation of events'.¹⁰ In this article, we adopt the terminology used by the ALRC. We describe misconduct falling within the system of conduct or pattern of behaviour concept as 'systematic' misconduct and use the term 'system of conduct' as a shorthand for this concept.

Part II of this article explores the evolution of the criminal law and its application to corporations in order to explain how criminal law developed with natural persons, rather than corporations, in mind. Part III builds on that historical analysis, and considers how reliance on principles of attribution under traditional

⁴ Crimes Legislation Amendment (Combatting Corporate Crime) Bill 2019 (Cth) sch 1 cl 8 (proposed s 70.5A of the *Criminal Code* (Cth) (n 1)). Such an offence was introduced in the United Kingdom in s 7 of the *Bribery Act 2010* (UK). See discussion in *ALRC Corporate Criminal Responsibility Report* (n 3) 296–321 [7.63]–[7.177].

⁵ For an analysis of failure to prevent offences in the context of corporate misconduct uncovered during recent Australian Royal Commissions: see Penny Crofts, 'Three Recent Royal Commissions: The Failure to Prevent Harms and Attributions of Organisational Liability' (2020) 42(4) *Sydney Law Review* 395.

⁶ *ALRC Corporate Criminal Responsibility Report* (n 3).

⁷ *Competition and Consumer Act 2010* (Cth) sch 2 ('*Australian Consumer Law*') s 21(4); *Australian Securities and Investments Commission Act 2001* (Cth) s 12CB ('*ASIC Act*').

⁸ See *ALRC Corporate Criminal Responsibility Report* (n 3) 270–96 [7.6]–[7.62].

⁹ *ALRC Corporate Criminal Responsibility Report* (n 3) 275 [7.21].

¹⁰ *Unique International College Pty Ltd v Australian Competition and Consumer Commission* (2018) 266 FCR 631, 654 [104] (Allsop CJ, Middleton and Mortimer JJ) ('*Unique v ACCC*').

models of corporate criminal responsibility has created the need for offences tailored specifically to corporations. It outlines some of the existing models of such offences, such as duty-based offences and failure to prevent offences. Part IV discusses the system of conduct or pattern of behaviour concept adopted by the ALRC as a foundation for framing system of conduct offences. Part V considers how such an offence model might be utilised beyond the context recommended by the ALRC. Part VI concludes.

II Evolution of the Criminal Law and its Application to Corporations

To understand the contemporary need for offences tailored to corporations, an understanding of two aspects of the historical context is necessary. First, criminal law's regulatory role has developed incrementally over time in response to specific developments both in society and in the nature of commerce. Second, the law of criminal responsibility developed with natural persons, rather than corporate entities, as its focus. The law developed methods and analogies to apply the criminal law to corporations while seeking to preserve the fundamental structure of criminal offences, including the need to prove both physical acts and states of mind.

A *The Historical Development of Criminal Law as a Regulatory Tool*

Today, the use of the criminal law as a regulatory tool is ubiquitous and expanding, at least as a matter of legislative enactment.¹¹ The regulatory utility of criminalisation arises due to its unique expressive power — its capacity to condemn and denounce serious misconduct.¹² This means that the criminal law is a 'regulatory tool for influencing behaviour', but also 'speaks with a distinctively moral voice'.¹³ This reflects the origins of the criminal law. Prior to the 19th century, criminal law was more closely confined to those criminal offences where there was a direct 'connection to ... serious moral wrongdoing: broadly speaking, offences against

¹¹ For example, the ALRC recently identified 3,117 offences across 25 Commonwealth statutes as potentially applicable to corporations: *ALRC Corporate Criminal Responsibility Report* (n 3) 74 [3.13]. These ranged significantly in terms of seriousness and in terms of overlap with civil penalty provisions: at 73–4 [3.12]–[3.13]. Furthermore, a number of new criminal offence provisions have been created in the implementation of the recommendations of the *Australian Securities and Investments Commission* ('ASIC') Enforcement Review: for a summary of the recommendations, see Treasury (Cth), *ASIC Enforcement Review Taskforce Report* (December 2017) xiv–xviii.

¹² Samuel Walpole, 'Criminal Responsibility as a Distinctive Form of Corporate Regulation' (2020) 35(2) *Australian Journal of Corporate Law* 235, 255–61; *ALRC Corporate Criminal Responsibility Report* (n 3) 195–9 [5.78]–[5.89].

¹³ AP Simester and Andreas von Hirsch, *Crimes, Harms and Wrongs: On the Principles of Criminalisation* (Hart Publishing, 2011) 4.

religion, against the state, against the person, or against property'.¹⁴ The focus was on moral wrongdoing by *individuals*.¹⁵

There was a 'significant expansion of regulatory criminal offences during the mid-nineteenth century' as part of the Industrial Revolution, as Horder has observed.¹⁶ Indeed, Horder suggests the use of the criminal law to regulate behaviour occurred long before then.¹⁷ Why did the criminal law expand into a tool of regulation? Why was regulation not left to the private law, like the regulation of trusts or contracts? The growth of the regulatory state in the 19th century remains a topic of debate among both lawyers and historians.¹⁸ Criminal prosecutions were not initially a common 'regulatory tool' in relation to white-collar crime.¹⁹ This was partly because prosecutions were mostly conducted privately, in contrast to 'current understandings of the role of the state in protecting citizens from harm, which defines the unique character of the criminal law, and helps to distinguish it from civil wrongs'.²⁰ Private prosecutions persisted due to a political philosophy of minimal state intervention in order to promote personal liberty.²¹

Although the criminal law had been used in England to regulate conduct since the Middle Ages, such use grew dramatically in the 19th century.²² This 'administrative revolution' transformed the enforcement of regulatory norms into something done centrally by the State, using the criminal law.²³ MacDonagh 'proposed a five stage model' for how this growth in regulation occurred:

[F]irst, public exposure of an intolerable social evil; secondly, legislation to deal with it, which due to inexperience was ineffective; thirdly, the introduction of more effective procedures of enforcement or detection, which continually revealed new problems; fourthly, recognition that occasional parliamentary legislation was inadequate and continuous regulation was required in the light of growing and changing experience; finally,

¹⁴ Jeremy Horder, 'Bureaucratic "Criminal" Law: Too Much of a Bad Thing?' in R A Duff et al (eds), *Criminalization: The Political Morality of the Criminal Law* (Oxford University Press, 2014) 101, 103. Horder, however, expresses scepticism of such accounts.

¹⁵ Nicola Lacey, 'Philosophical Foundations of the Common Law: Social Not Metaphysical' [2000] *SSRN Electronic Journal* 24.

¹⁶ Horder (n 14) 102–5.

¹⁷ *Ibid* 105.

¹⁸ See, eg, Horder (n 14); Oliver MacDonagh, 'The Nineteenth-Century Revolution in Government: A Reappraisal' (1958) 1(1) *The Historical Journal* 52; Henry Parris, 'The Nineteenth-Century Revolution in Government: A Reappraisal Reappraised' (1960) 3(1) *The Historical Journal* 17.

¹⁹ James Taylor, 'Company Fraud in Victorian Britain: The Royal British Bank Scandal of 1856' (2007) 122(497) *English Historical Review* 700, 723. See also James Taylor, 'White-Collar Crime and the Law in Nineteenth-Century Britain' (2018) 60(3) *Business History* 343, 347; Nuno Garoupa, Anthony Ogus and Andrew Sanders, 'The Investigation and Prosecution of Regulatory Offences: Is There an Economic Case for Integration?' (2011) 70(1) *Cambridge Law Journal* 229, 230–4.

²⁰ Sarah Wilson, 'Law, Morality and Regulation' (2006) 46(6) *The British Journal of Criminology* 1073, 1079.

²¹ Paul Rock, 'Victims, Prosecutors and the State in Nineteenth Century England and Wales' (2004) 4(4) *Criminal Justice* 331, 340. See also John Braithwaite, 'The New Regulatory State and the Transformation of Criminology' (2000) 40(2) *British Journal of Criminology* 222, 223.

²² Garoupa, Ogus and Sanders (n 19) 237.

²³ *Ibid* 237–8.

discretionary initiative was given to executive officers to deal with problems as they were continually revealed.²⁴

Other scholars have argued that the key driver of the growth in regulation by the State — and through the criminal law — was the influence of utilitarianism during this period, which ‘led to considerable extensions both of *laissez-faire* and of State intervention simultaneously’.²⁵ Criminal law came to embrace dual functions as both a ‘moral and ... retributive system’ and as an enforcement mechanism with a ‘regulatory, instrumental or utilitarian aspect’.²⁶ The expanded range of regulatory offences

were a product of the expanding functions of the modern administrative state, for which the criminal law became an increasingly important tool for regulating the areas of social life born of industrialisation and urbanisation from the early nineteenth century onwards.²⁷

This expansion of the regulatory role of the criminal law ‘underlines a more general transformation in the legal order’ during this period, as criminal justice evolved into ‘a matter of administration and security’.²⁸ As Jackson J explained in *Morisette v United States*:

The industrial revolution multiplied the number of workmen exposed to injury from increasingly powerful and complex mechanisms, driven by freshly discovered sources of energy, requiring higher precautions by employers. Traffic of velocities, volumes and varieties unheard of came to subject the wayfarer to intolerable casualty risks if owners and drivers were not to observe new cares and uniformities of conduct. Congestion of cities and crowding of quarters called for health and welfare regulations undreamed of in simpler times. Wide distribution of goods became an instrument of wide distribution of harm when those who dispersed food, drink, drugs, and even securities, did not comply with reasonable standards of quality, integrity, disclosure and care. Such dangers have engendered increasingly numerous and detailed regulations which heighten the duties of those in control of particular industries, trades, properties or activities that affect public health, safety or welfare.²⁹

B *Application of the Criminal Law to Corporations*

Even as criminal law’s regulatory function expanded, it continued to focus on crimes by natural persons. For much of the history of the common law, corporations could

²⁴ As summarised in Pat Thane, ‘Government and Society in England and Wales, 1750–1914’ in FML Thompson (ed), *The Cambridge Social History of Britain, 1750–1950* (Cambridge University Press, 1st ed, 1990) 1, 19.

²⁵ Parris (n 18) 35.

²⁶ Nicola Lacey, Celia Wells and Oliver Quick, *Reconstructing Criminal Law: Text and Materials* (Cambridge University Press, 4th ed, 2010) 6.

²⁷ *Ibid* 7.

²⁸ Lindsay Farmer, ‘The Obsession with Definition: The Nature of Crime and Critical Legal Theory’ (1996) 5(1) *Social & Legal Studies* 57, 65–6.

²⁹ *Morisette v United States*, 342 US 246, 253–4 (1952).

not be guilty of a criminal offence.³⁰ The common law recognised the corporation as a juristic entity, but did not confer on it the capacity to be criminally responsible.³¹ The common law at this time lacked any model of corporate fault, arguably due to the absence of any appropriate principles of attribution to capture such a concept.³²

Eventually the common law recognised that a corporation could be the subject of the criminal law, but it lacked principles for constructing a corporate ‘state of mind’.³³ The criminal law was extended to corporations initially for particular types of criminal offences: nuisance, breach of statutory duties, and public welfare offences.³⁴ This was followed by offences of misfeasance that did not require proof of a mental element.³⁵ By the 1930s, courts in Australia, England, New Zealand, and the United States had held that a corporation could also be guilty of an offence that required proof of a mental element.³⁶ Consequently, corporate criminal liability — of some form — has been an established part of the criminal law in common law jurisdictions for over a century. The approach to corporate criminal responsibility based on attribution that developed was an evolution of the criminal law’s traditional focus on individuals. As a result, corporate criminal responsibility may seem to reflect ‘a form of anthropomorphism’.³⁷

The present position under Commonwealth criminal law in Australia is that all criminal offences apply generally to corporations.³⁸ Section 12.1(2) of the *Criminal Code Act 1995* (Cth) (*‘Criminal Code (Cth)’*) provides that a ‘body corporate may be found guilty of any offence, including one punishable by imprisonment’. Section 12.1(1) states that:

This Code applies to bodies corporate in the same way as it applies to individuals. It so applies with such modifications as are set out in [Part 2.5] and with such other modifications as are made necessary by the fact that criminal liability is being imposed on bodies corporate rather than individuals.

³⁰ *The Abbot of St Benet’s v Mayor of Norwich* (1481) YB 21 Edw IV 7, 12, 27, 67; *Case of Sutton’s Hospital* (1612) 10 Co Rep 23a; 77 ER 960, 973; *Anon* (1706) 88 ER 1518; William Blackstone, *Commentaries on the Law of England* (Clarendon Press, 1765) vol 1, 464–5.

³¹ Walpole (n 12) 236, 248–55.

³² *Ibid* 248.

³³ See discussion *ibid* 241–8. Conversely, the development of corporate criminal liability in Civilian jurisdictions was slower and, in some jurisdictions, has not occurred: Guy Stessens, ‘Corporate Criminal Liability: A Comparative Perspective’ (1994) 43(3) *International & Comparative Law Quarterly* 493.

³⁴ LH Leigh, *The Criminal Liability of Corporations in English Law* (Weidenfeld and Nicholson, 1969) 15.

³⁵ *R v Great North of England Railway Company* (1846) 9 QB 315.

³⁶ See, eg, *New York Central and Hudson River Railroad Company v United States*, 212 US 481 (1909) (*‘New York Central’*); *Mousell Bros Ltd v London and North Western Railway Co* [1917] 2 KB 836 (*‘Mousell Bros’*); *R v Australasian Films Ltd* (1921) 29 CLR 195 (*‘Australasian Films’*); *OF Nelson and Co Ltd v Police* (1931) 4 NZ Police Reports 248. The historical development in Canada was slightly more complex: see, eg, *R v Fane Robinson Ltd* [1941] 3 DLR 409; *Canadian Dredge & Dock Co v The Queen* [1985] 1 SCR 662 (*‘Canadian Dredge & Dock’*). See also LH Leigh, ‘The Criminal Liability of Corporations and Other Groups’ (1977) 9(2) *Ottawa Law Review* 247, 249–51. For a discussion of how corporate criminal responsibility developed over the course of the first half of the 20th century in several common law jurisdictions, refer to the judgment of Estey J in *Canadian Dredge & Dock*: at 674–91.

³⁷ Lacey (n 15) 24.

³⁸ *Criminal Code* (Cth) (n 1) s 12.1; *Acts Interpretation Act 1901* (Cth) s 2C(1); *Crimes Act 1914* (Cth) s 4B(1).

Section 12.1 essentially renders the question whether a corporation can commit a criminal offence one of statutory interpretation, with a presumption that corporations can be guilty of the same offences as a natural person. This approach is consistent with other Commonwealth statutes. For example, s 2C(1) of the *Acts Interpretation Act 1901* (Cth) provides that '[i]n any act, expressions used to denote persons generally ... include a body politic or corporate as well as an individual.' Similarly to s 12.1 of the *Criminal Code* (Cth), s 4B(1) of the *Crimes Act 1914* (Cth) provides that a 'provision of a law of the Commonwealth relating to indictable offences or summary offences shall, unless the contrary intention appears, be deemed to refer to bodies corporate as well as natural persons'. As Allsop P opined in *Presidential Security Services of Australia Pty Ltd v Brille*, given the developments in principles of attribution of criminal responsibility to corporations, '[t]he identity of the offences that might be considered, from their very nature, not capable of commission by a company ... must now be narrow.'³⁹

The general position under such statutory provisions is that criminal offences are as capable of commission by a corporation as they are by a natural person. However, the nature of a corporate defendant and the nature of the offence itself may mean that, in practice, a corporation cannot commit certain criminal offences. Therefore, there is, in fact, a recognition by the legislature in these statutory provisions of the necessity, in relation to certain offences, to treat corporations differently. The starting position, however, is that all criminal offences are generally considered to be applicable to corporations.⁴⁰

Despite the now well-established existence of corporate criminal responsibility, the low incidence of criminal prosecutions of corporations⁴¹ shows that corporations are not, in reality, treated the same as individuals insofar as the criminal law is concerned. Discussions by the ALRC with regulators also evinced a greater willingness to pursue the relevant individuals criminally, rather than the corporation.⁴²

C *Mens Rea and its Imperfect Fit with the Criminal Liability of Corporations*

Generally, punishment for a criminal offence is only justified where a person is morally blameworthy. Traditionally, this requires proof of both physical and mental elements,⁴³ with the mental element required to establish criminal responsibility being 'mens rea or guilty mind'.⁴⁴ As Brennan J observed in *He Kaw Teh v The*

³⁹ *Presidential Security Services of Australia Pty Ltd v Brille* (2008) 73 NSWLR 241, 248 [20]–[21]. Some examples of offences that may be uniquely human include bigamy and perjury. Relevant to that appeal was s 10(1) of the *Criminal Procedure Act 1986* (NSW) which is similar in effect to the Commonwealth statutory provisions set out in the text.

⁴⁰ *Criminal Code* (Cth) (n 1) s 12.1; *Acts Interpretation Act 1901* (Cth) (n 38) s 2C(1); *Crimes Act 1914* (Cth) (n 38) s 4B(1).

⁴¹ See *ALRC Corporate Criminal Responsibility Report* (n 3) 96–117 [3.69]–[3.110].

⁴² Such discussions occurred in the course of consultations undertaken during the ALRC's Corporate Criminal Responsibility Inquiry.

⁴³ Thomas Weigend, 'Subjective Elements of Criminal Liability' in Markus D Dubber and Tatjana Hörnle (eds), *The Oxford Handbook of Criminal Law* (Oxford University Press, 2014) 490, 490.

⁴⁴ *He Kaw Teh v The Queen* (1985) 157 CLR 523, 565 (Brennan J) ('*He Kaw Teh*').

Queen, the ‘requirement of mens rea avoids ... “the public scandal of convicting on a serious charge persons who are in no way blameworthy”’.⁴⁵ Consequently, there is a presumption that a statutory offence requires proof of mens rea.⁴⁶ Offences of strict and absolute liability are exceptions to this principle.⁴⁷ While many lower-level regulatory offences involve strict or absolute liability, for most serious criminal offences it remains the case that proof of a ‘guilty mind’ is necessary to establish the criminal offence. The need to prove mens rea has been at the core of difficulties in applying the criminal law to corporate defendants. It is why corporations remain ‘penumbral subjects of criminal law’.⁴⁸

The doctrine of mens rea can be traced back to the medieval period — even before its express recognition as an element of criminal responsibility.⁴⁹ Papp Kamali has explored medieval jurors’ focus on ‘intentionality’ and the reception of the idea that ‘culpability depends upon the presence of *mens rea*, or guilty mind’ into medieval English criminal law through canonist influences.⁵⁰ The importance of ‘mind’ stemmed from the religious beliefs of the time.⁵¹ These principles developed into the modern doctrine of mens rea.⁵² One can observe an inherently human aspect to this conception of culpability that does not easily translate to a corporate defendant. The requirement to establish mens rea was the real sticking point for the general acceptance of corporate criminal responsibility, outside of specific categories of offences⁵³ — ‘[t]he most formidable impediment ... [to corporate criminal responsibility] ... was the doctrine that a corporation does not have a mind and hence is incapable of mens rea.’⁵⁴ The 1909 edition of *Halsbury’s Laws of England* stated that:

By the general principles of the criminal law, if a matter is made a criminal offence it is essential that there should be something in the nature of *mens rea*, and therefore, in ordinary cases, a corporation aggregate cannot be guilty of a criminal offence.⁵⁵

The law was required to construct models of corporate fault through which the requirements of criminal responsibility relevant to individuals could be translated to corporations.⁵⁶ Early attempts adopted vicarious liability as the method of

⁴⁵ Ibid, quoting *Sweet v Parsley* [1970] AC 132, 150 (Lord Reid).

⁴⁶ *He Kaw Teh* (n 44) 565–6 (Brennan J).

⁴⁷ Brennan J explained that such offences are justified where ‘the purpose of the [offence] is not merely to deter a person from engaging in prohibited conduct but to compel him to take preventive measures to avoid the possibility that, without deliberate conduct on his part, the external elements of the offence might occur’: ibid 567.

⁴⁸ Lacey (n 15) 18. Lacey observed: ‘[T]he subject of modern systems of criminal law in and beyond the common law world is generally assumed to be a human individual. The elaborated notions of conduct and responsibility ... have been worked out in relation to assumptions about individual human beings’: at 19.

⁴⁹ Elizabeth Papp Kamali, *Felony and the Guilty Mind in Medieval England* (Cambridge University Press, 2019) 305–8.

⁵⁰ Ibid 2–3.

⁵¹ Ibid 2, 10–11.

⁵² Ibid 4–11, 305.

⁵³ See Lacey (n 15) 46.

⁵⁴ *R v JJ Beamish Construction Co Ltd* [1966] 2 OR 867, 883–4 (Jessup J).

⁵⁵ *Halsbury’s Laws of England* (Butterworths, 1st ed, 1909) vol 8, 390 [858] quoted in *Canadian Dredge & Dock* (n 36) 676.

⁵⁶ See Walpole (n 12).

attribution.⁵⁷ Since then, English and Australian law has developed more sophisticated mechanisms of attribution,⁵⁸ such as identification theory and, later, models of organisational liability that seek to ascribe particular conduct and mental states to the corporation itself, rather than merely holding the corporation vicariously liable for the conduct of its agents.

Identification theory, in its original form, involved a search for the ‘directing mind and will’ of the corporation, whose acts and mental states were to be attributed to the corporation.⁵⁹ As such, identification theory considers the mental states of the individuals who direct the corporation to determine the corporation’s criminality. In *Meridian Global Funds Management Asia Ltd v Securities Commission*, which reflects the current common law approach in Australia,⁶⁰ the identification approach was refined so that the question became who was the company for the particular offence, based on statutory interpretation.⁶¹ Statutory innovations have also sought to widen the identification doctrine. The prevailing attribution method under Commonwealth criminal law, which the ALRC termed the ‘TPA Model’,⁶² does so by deeming the states of mind and conduct of a director, employee, or agent to be those of the company.⁶³ We discuss the limitations of these approaches further below. Conversely, organisational models of liability, like that in pt 2.5 of the *Criminal Code* (Cth), seek to construct a model of corporate fault unique to the corporation itself.⁶⁴

⁵⁷ See, eg, *Mousell Bros* (n 36) 845–6 (Viscount Reading CJ, Ridley and Atkin JJ); *New York Central* (n 36) 495 (Day J); *Australasian Films Ltd* (n 36) 217 (Knox CJ, Gavan Duffy and Rich JJ). In regards to the debate about whether vicarious liability involves derivative liability for an agent’s conduct or an agent’s own liability, see *Pioneer Mortgage Services Pty Ltd v Columbus Capital Pty Ltd* (2016) 250 FCR 136, 147–9 [48]–[58] (Davies, Gleeson and Edelman JJ).

⁵⁸ United States jurisdictions (including federal law) continue predominantly to use vicarious liability — *respondeat superior* — as their method of attribution pursuant to *New York Central* (n 36). Some state decisions have adopted an approach more in line with identification theory: see, eg, *People v Canadian Fur Trappers Corp*, 248 NY 159, 163, 169 (NY, 1928); *Idaho v Adjustment Department Credit Bureau*, 483 P 2d 687, 691 (Idaho, 1971); *St Johnsbury Trucking Co v United States*, 220 F 2d 393 (1st Cir, 1955); *Louisiana v Chapman Dodge Center Inc*, 428 So 2d 413 (La, 1983); *Commonwealth v Beneficial Finance Co*, 360 Mass 188 (Mass, 1971). See discussion in *Canadian Dredge & Dock* (n 36) 686–8; Eliezer Lederman, ‘Criminal Law, Perpetrator and Corporation: Rethinking a Complex Triangle’ (1985) 76(2) *The Journal of Criminal Law and Criminology* 285, 288–93.

⁵⁹ See *Lennard’s Carrying Co v Asiatic Petroleum* [1915] AC 705, 713–14 (Viscount Haldane LC); *HL Bolton (Engineering) Co Ltd v TJ Graham & Sons Ltd* [1957] 1 QB 159, 172 (Denning LJ); *Tesco Supermarkets Ltd v Natrass* [1972] AC 153, 170–2 (Lord Reid) (*‘Tesco’*).

⁶⁰ In both a criminal and civil context, see, eg: *ABC Developmental Learning Centres Pty Ltd v Joanne Wallace* (2006) 161 A Crim R 250; *Director of Public Prosecutions Reference No 1 of 1996* [1998] 3 VR 352; *Australian Securities and Investments Commission v Westpac Banking Corporation (No 2)* (2018) 266 FCR 147 (*‘ASIC v Westpac (No 2)’*).

⁶¹ *Meridian Global Funds Management Asia Ltd v Securities Commission* [1995] 2 AC 500, 507 (Lord Hoffmann) (*‘Meridian’*).

⁶² Due to its origin in s 84 of the then *Trade Practices Act 1974* (Cth) (now the *Competition and Consumer Act 2010* (Cth)). See *ALRC Corporate Criminal Responsibility Report* (n 3) 90–3 [3.57]–[3.65].

⁶³ The defining features of the TPA Model are discussed in detail in *ALRC Corporate Criminal Responsibility Report* (n 3) 250–6 [6.123]–[6.150].

⁶⁴ See *ALRC Corporate Criminal Responsibility Report* (n 3) 147–8 [4.73]–[4.79].

III Alternative Approaches to Corporate Criminal Liability

A Overcoming Limitations of Traditional Approaches through Tailored Offences

There is increasing interest among policymakers in models of corporate criminal liability that better reflect the nature of the corporation and the reality of corporate action than traditional approaches relying on attribution. This is because alternative models can overcome the limitations associated with establishing criminal responsibility through traditional approaches based on attribution principles. Attribution is inherently artificial. Whether the method used is the common law identification theory (as refined in *Meridian*)⁶⁵ or the TPA Model, the central limitation of attribution theory is that it seeks to locate corporate action and states of mind — themselves constructs — in a particular human individual or individuals. As Lord Hoffmann said in *Meridian*, the question is ‘Whose act (or knowledge, or state of mind) was for this purpose intended to count as the act etc. of the company?’⁶⁶ Both *Meridian* and the TPA Model merely widen the scope of actors who may be identified with the company beyond the ‘directing mind and will’ model of the mid-20th century.⁶⁷

The central limitation of all of these approaches, which are tied to particular individuals, is that they do not have regard to the complexity of a modern corporation. As Bant has explained:

With some minor exceptions, [these attribution methods] ultimately require identification of one human repository of the requisite knowledge or state of mind. For this reason, they are routinely confounded by the dispersed lines of authority, staff turnover, knowledge silos and fragmented task responsibility that characterise many corporate business models. Indeed ... it is arguable that these attribution rules actually encourage corporate structures that disperse knowledge and responsibility.⁶⁸

The innovative attribution method enacted in pt 2.5 of the *Criminal Code* (Cth) goes some way to ameliorating these limitations, through enabling a prosecutor to prove fault through the presence or absence of a corporate culture.⁶⁹ While corporate culture looks effective in theory, particularly as it embraces notions of ‘organisational liability’,⁷⁰ in reality most statutory regimes exclude it.⁷¹ As a result, there is a lack of experience and practical guidance as to how it might apply in

⁶⁵ *Meridian* (n 61).

⁶⁶ Ibid 507.

⁶⁷ See *Tesco* (n 59).

⁶⁸ Elise Bant, Submission No 21 to ALRC, *Corporate Criminal Responsibility: Discussion Paper* (28 January 2020) 2.

⁶⁹ *Criminal Code* (Cth) (n 1) s 12.3.

⁷⁰ Models based on ‘organisational liability’ seek to construct a ‘holistic model of corporate fault’ that has regard to ‘how the corporation operates, as a collection of systems and relationships’: ALRC *Corporate Criminal Responsibility Report* (n 3) 147 [4.73]–[4.74].

⁷¹ ALRC *Corporate Criminal Responsibility Report* (n 3) 90–2 [3.57]–[3.63]. For example, s 769A of the *Corporations Act 2001* (Cth) (‘*Corporations Act*’) expressly excludes pt 2.5 from applying to ch 7 of the *Corporations Act* and s 12GH(6) of the *ASIC Act* (n 7) expressly excludes its application to the provisions of pt 2 div 2 of the *ASIC Act* (n 7).

practice, and an absence of concrete evidence of its efficacy at securing corporate accountability.

Ultimately, this suggests that, in certain contexts, alternative approaches to corporate criminal liability may be justified as a means of overcoming the limitations of traditional approaches. As Crofts has cogently argued, the fact that '[c]riminal legal doctrine has failed to develop a coherent, persuasive and effective means of attributing responsibility for harms caused by large organisations' shows a need for novel criminal offences.⁷² Consequently, '[t]he failure of the criminal justice system to respond to systemic failures of large organisations requires us to think imaginatively and broadly about organisational culpability.'⁷³ Such an approach is already underway. In the remainder of this Part of the article, we describe two existing types of offences tailored to corporations: duty-based offences, and failure to prevent offences. We outline these in order to situate consideration of a new type of tailored offence based on systems of conduct within the existing policy trend toward tailored offences.

B *Duty-based Offences*

Duty-based offences are not a new invention.⁷⁴ They were 'the first inroad upon the proposition ... [that] ... "A corporation is not indictable..."',⁷⁵ and pre-date the development of generalised principles of corporate attribution.⁷⁶ There is, however, an 'increasing preference for such offences in legislation relevant to industries involving high proportions of corporate actors'.⁷⁷ The basic principle behind a duty-based offence is that the corporation *itself* is held liable for its *own* breach of a particular duty imposed on it by statute.⁷⁸ It follows that '[t]here is no need to find someone — in the case of a company, the "brains" and not merely the "hands" — for whose acts the person with the duty can be held liable'.⁷⁹ As the Victorian Court of Appeal recently stated in relation to a particular duty-based offence: 'it is tolerably clear that the rules of attribution do not apply'.⁸⁰ The ALRC observed that the key advantage of a duty-based offence in the corporate context is that, properly framed,

⁷² Crofts (n 5) 397.

⁷³ Ibid 400.

⁷⁴ In *R v Birmingham and Gloucester Railway Company* it was held that a corporation could be guilty of non-feasance in breach of a statutory duty requiring the corporation to do something: *R v Birmingham and Gloucester Railway Company* (1842) 3 QB 223. Then, in *R v Great North of England Railway Company* (n 35) it was held, following the earlier case, that the criminal liability of a corporation extended to misfeasance in the exercise of powers conferred under a statute. For a summary of these earlier cases, see CRN Winn, 'The Criminal Responsibility of Corporations' (1929) 3(3) *Cambridge Law Journal* 398.

⁷⁵ Winn (n 74) 399.

⁷⁶ Meaghan Wilkinson, 'Corporate Criminal Liability: The Move towards Recognising Genuine Corporate Fault' (2003) 9 *Canterbury Law Review* 142, 143–8; Stessens (n 33) 496, cited in *ALRC Corporate Criminal Responsibility Report* (n 3) 321 [7.178].

⁷⁷ *ALRC Corporate Criminal Responsibility Report* (n 3) 326 [7.193].

⁷⁸ Winn (n 74) 398.

⁷⁹ Sir John Smith, 'Health and Safety at Work' [1995] *Criminal Law Review* 654, 655.

⁸⁰ *Director of Public Prosecutions (Vic) v JCS Fabrications Pty Ltd* [2019] VSCA 50, 19 [39] (Whelan AP, Priest and McLeish JJA).

it can avoid the limitations of models of corporate liability based on principles of attribution.⁸¹

This type of offence is commonly used in a workplace health and safety context. Under the harmonised approach to workplace health and safety laws in Australia,⁸² it is an offence⁸³ for ‘a person conducting a business or undertaking’⁸⁴ to fail to comply with a number of broad duties of care imposed by the legislation.⁸⁵ The *Corporate Manslaughter and Corporate Homicide Act 2007* (UK) contains a duty-based offence expressly focused on a corporation. It establishes a specific offence for a corporation whose ‘activities are managed or organised’ in such a way as to cause a person’s death and where the way in which those activities are managed or organised ‘amounts to a gross breach of a relevant duty of care owed by the organisation to the deceased’.⁸⁶ Outside of the workplace health and safety context, duty-based offences have been enacted in relation to the regulation of heavy vehicles and in certain environmental protection legislation.⁸⁷

C Failure to Prevent Offences

Another type of offence that is increasing in prevalence is offences that adopt the ‘failure to prevent’ model. The failure to prevent model is based on an organisational liability framework. Offences are framed in terms of the corporation itself committing an offence due to its failure to prevent the commission of a predicate offence by one of its associates.⁸⁸ Failure to prevent offences were first enacted in respect of foreign bribery in the *Bribery Act 2010* (UK)⁸⁹ and have since been expanded to apply to tax avoidance.⁹⁰ The United Kingdom has also given consideration to expanding the model to ‘economic crime’ more generally, and to corporate human rights violations.⁹¹ In Australia, an offence of failing to prevent foreign bribery is contained in the Crimes Legislation Amendment (Combatting

⁸¹ *ALRC Corporate Criminal Responsibility Report* (n 3) 322–3 [7.180]. Of course, where a particular duty-based offence requires proof of recklessness, it will be impossible to avoid the use of principles of attribution in relation to a corporate defendant: at 325 [7.190].

⁸² Based on the *WHS Act* (Cth) (n 3). Only Victoria and Western Australia retain their own legislative framework, although these regimes also contain duty-based offences: *ALRC Corporate Criminal Responsibility Report* (n 3) 323 [7.184].

⁸³ *WHS Act* (Cth) (n 3) pt 2 div 5.

⁸⁴ *Ibid* s 5. It follows that these duty-based offences are not expressly restricted to corporate defendants. However, such offences are ‘often and effectively prosecuted against corporations’ and ‘[t]his was clearly the legislative intent behind the drafting of such offences’: *ALRC Corporate Criminal Responsibility Report* (n 3) [7.191]. This, coupled with their aptness for application to a corporation means they can properly be considered to be offences tailored to corporate (in)action.

⁸⁵ Relevantly to a corporate defendant, see *WHS Act* (Cth) (n 3) pt 2 divs 2–3.

⁸⁶ *Corporate Manslaughter and Corporate Homicide Act 2007* (UK) s 1.

⁸⁷ See *ALRC Corporate Criminal Responsibility Report* (n 3) 326 [7.193]–[7.194], referring to the *Heavy Vehicle National Law* and the *Environmental Protection Amendment Act 2018* (Vic) s 25(1).

⁸⁸ *Ibid* [7.63].

⁸⁹ *Bribery Act 2010* (UK) s 7.

⁹⁰ *Criminal Finances Act 2017* (UK) ss 45–6.

⁹¹ Liz Campbell, ‘Corporate Liability and the Criminalisation of Failure’ (2018) 12(2) *Law and Financial Markets Review* 57, 61.

Corporate Crime) Bill 2019 (Cth), which remains before the Australian Parliament.⁹² The ALRC also recommended '[t]he Australian Government ... consider applying the failure to prevent [model] ... to other Commonwealth offences that might arise in the context of transnational business'.⁹³

The failure to prevent model involves

a standalone offence under which a corporation can be convicted of failing to prevent the commission of a stipulated primary offence ... by one of its 'associates'. A defence of appropriate or reasonable measures (or due diligence) allows a corporation to show that it lacks organisational culpability if it can prove that targeted policies and procedures were in place to prevent the offence.⁹⁴

This type of offence avoids the need to prove the fault of the corporation *itself* through principles of attribution.⁹⁵ Instead, the conceptualisation of corporate fault constructed under the failure to prevent model holds the corporation as an entity liable for failing to prevent the predicate offence.⁹⁶ The corporation is called on to prove an absence of corporate fault through establishing a defence of 'adequate procedures',⁹⁷ 'reasonable precautions',⁹⁸ 'reasonable measures',⁹⁹ or 'due diligence'¹⁰⁰ (depending on how the defence is framed in the statute). If a corporation can prove the defence on the balance of probabilities, then it cannot be said to be organisationally liable for the predicate offence committed by its associate. Campbell has argued that failure to prevent liability operates 'as a preventative device and a mechanism to influence behaviour, rather than something that operates primarily in reactive mode'.¹⁰¹ The 'adequate procedures' defence gives a corporation an incentive to improve its business practices, systems, and procedures. On a more theoretical level, Crofts has argued that failure to prevent offences are justified on the basis of the 'harmful consequences' of the failure 'and the blameworthiness of the failure/s'.¹⁰²

⁹² Crimes Legislation Amendment (Combatting Corporate Crime) Bill 2019 (Cth) sch 1 cl 8 (proposed s 70.5A of the *Criminal Code* (Cth) (n 1)).

⁹³ *ALRC Corporate Criminal Responsibility Report* (n 3) 18 recommendation 19.

⁹⁴ *Ibid* 297 [7.67].

⁹⁵ Although it remains necessary to prove the mental elements of the predicate offence committed by the associate that the corporation is alleged to have failed to prevent. There is no requirement, however, to attribute these fault elements to the corporation, which is where one of the advantages of failure to prevent offences arises.

⁹⁶ Crofts has argued, based on recent systemic failures by organisations in Australia, that the requirement to prove the predicate offence should be removed and instead criminal liability should result where there has been a 'systemic failure to prevent breach of [a] legal duty of care': Crofts (n 5) 397. Such an approach would appear to blend aspects of duty-based and failure to prevent offences. See also 409–11.

⁹⁷ See, eg, *ALRC Corporate Criminal Responsibility Report* (n 3) 304–6 [7.96]–[7.109]; 312–15 [7.134]–[7.147].

⁹⁸ This was the formulation ultimately recommended by the ALRC in the context of amendments to the methods of attribution applicable under Commonwealth criminal law: *ALRC Corporate Criminal Responsibility Report* (n 3) 259–66 [6.161]–[6.192].

⁹⁹ See, eg, *Criminal Code*, RSC 1985, c C-46, s 22.2(c).

¹⁰⁰ See, eg, *Criminal Code* (Cth) (n 1) s 12.3(3).

¹⁰¹ Campbell (n 91) 59.

¹⁰² Crofts (n 5) 418–22.

IV Addressing Systematic Misconduct by Corporate Entities

In recent years, it has become apparent that there is a need to address deficient corporate systems and processes that result in corporate misconduct, whether civil or criminal. An offence targeted to respond to systematic misconduct in a corporate context could help to address this. This Part discusses the need to respond to corporate systems and processes that contravene the law, and how, as the ALRC suggested, the system of conduct or pattern of behaviour concept used in existing civil regulatory provisions could form the foundation for a system of conduct offence. It concludes by discussing the particular type of system of conduct offence recommended by the ALRC.

A *Contemporary Corporate Action and Responding to Unlawful Corporate Systems and Processes*

Corporations both large and small are more than the product of the people they engage to act on their behalf as employees, contractors and agents. The value of the corporation is tied up in the policies and procedures, the ways of working, and norms of behaviour that result in the delivery of products or services — whether for profit or other purpose — in a way that is not possible through the mere agglomeration of people. In accounting terms, the value of a corporation is not just in tangible assets but also in the intangible assets such as goodwill.¹⁰³ The success of McDonald's as a quick service restaurant is as much a story of the 'McDonald's system' as the food served.¹⁰⁴

Just as a corporation's systems and procedures may serve its goals and benefit its consumers and shareholders, those very systems may result in great harm: maintenance systems may not be applied sufficiently to prevent oil being discharged from an offshore platform and polluting the marine environment; computer programs may produce insurance quotes that breach anti-discrimination laws; or, computer systems may send out invoices for financial services never delivered if human oversight is insufficient.¹⁰⁵ There is a need to grapple with the collective nature of the corporation that combines human endeavour with capital and technology when conceptualising what is criminal behaviour by a contemporary corporation.

In this way, the approach of looking to whether a single human within the corporation holds the requisite state of mind, whether as an employee under the TPA Model or as an individual who can be identified with the corporation at common

¹⁰³ See, eg, Australian Accounting Standards Board ('AASB'), *AASB 138 Intangible Assets* (F2021C00883, Compilation no 4, 30 June 2021).

¹⁰⁴ Mary-Angie Salva-Ramirez, 'McDonalds: A Prime Example of Corporate Culture' (1995–6) 40(4) *Public Relations Quarterly* 30, 30.

¹⁰⁵ See, eg, US Chemical Safety and Hazard Investigation Board, *Investigation Report, Explosion and Fire at the Macondo Well* (5 June 2014); Jeannie Marie Paterson and Yvette Maker, 'Why Does Artificial Intelligence Discriminate?' *Pursuit* (Web Page, 24 October 2018) <<https://pursuit.unimelb.edu.au/articles/why-does-artificial-intelligence-discriminate>>; Royal Commission into Misconduct in the Banking, Superannuation and Financial Services Industry (Final Report, February 2019) vol 1.

law, fundamentally misunderstands how modern, large corporations act in practice.¹⁰⁶ A corporation should not be absolved of liability simply because a bribe, paid to facilitate sales by a corporation in a foreign country, was made without the awareness of senior corporate officers in the company's headquarters. Those payments are the product of the systems, norms of conduct, and processes put in place (or not put in place) by the corporation. This is the same reasoning that underpins the enactment of failure to prevent offences in the context of foreign bribery.¹⁰⁷

The counterargument here is that mistakes or omissions by a corporation, even if occurring through systems or patterns of behaviour, are not equivalent to the positive states of mind generally required to establish many offences.¹⁰⁸ As a result, there is a risk of divergence between the criminal standard applied to humans and corporations.

This argument has some superficial attraction, particularly when examined through the lens of current tests of attribution. However, it is possible to consider systems of conduct or patterns of behaviour as indicia of corporate states of mind, or at least a reasonable alternative, in particular contexts.¹⁰⁹ Systems and patterns are something more than one-off mistakes; they reflect the way a corporation routinely acts in practice. In reality, attribution methods are themselves merely principles for constructing an artificial corporate state of mind. As Bant has previously explained, albeit in the civil context, '[t]he legitimate objective here is to regulate behaviour that has an inherently harmful tendency'.¹¹⁰ It may be appropriate 'to take seriously the original position taken by the courts, which was that as artificial entities, corporations lacked "minds," and instead focus on the objective quality of the conduct of corporations'.¹¹¹ As corporate conduct is 'unlikely to be unintentional in any relevant sense', it is more appropriate to 'focus ... on the quality of conduct, not on some artificial mental state'.¹¹² In addition, as noted above, while the criminal law has tried to apply offences in the same way to legal persons as it does to natural persons, the relevant statutes acknowledge that this will not always be possible.¹¹³ Similarly, Crofts has suggested that:

¹⁰⁶ *ALRC Corporate Criminal Responsibility Report* (n 3) 230 [6.42]. See also *Australian Securities and Investments Commission v King* (2020) 94 ALJR 293, 311 [92]–[93] (Nettle and Gordon JJ).

¹⁰⁷ See, eg, Steven Montagu-Cairns, 'Corporate Criminal Liability and the Failure to Prevent Offence: An Argument for the Adoption of an Omissions-Based Offence in AML' in Katie Benson, Colin King and Clive Walker (eds), *Assets, Crimes and the State: Innovation in 21st Century Legal Responses* (Routledge, 2020) 185.

¹⁰⁸ See Law Council of Australia Submission No 27 to ALRC, *Corporate Criminal Responsibility: Discussion Paper* (31 January 2020) 19 [70]. See also TA Game SC and Justice David Hammerschlag, Submission No 17 to ALRC, *Corporate Criminal Responsibility: Discussion Paper* (20 December 2019) 2–5 [8]–[24].

¹⁰⁹ See Elise Bant, 'Culpable Corporate Minds' (2021) 48(2) *University of Western Australia Law Review* 352. See also Elise Bant and Jeannie Marie Paterson, 'Systems of Misconduct: Corporate Culpability and Statutory Unconscionability' (2021) 15(1) *Journal of Equity* 63.

¹¹⁰ Bant (n 68) 6.

¹¹¹ *Ibid.*

¹¹² *Ibid* 7.

¹¹³ As set out in s 12.1 of the *Criminal Code* (Cth) (n 1), the Code is to be applied 'with such other modifications as are made necessary by the fact that criminal liability is being imposed on bodies corporate rather than individuals'.

there needs to be engagement with the types of harms most likely to be caused by large organisations, and the reasons why these harms come about. Whilst there are occasions where large organisations may actively choose to breach the law, it is more likely that breaches of the law are due to systemic failings on the part of the organisation. These systemic failings are culpable.¹¹⁴

We consider that systems-based liability should nonetheless be restricted to particular contexts where it is considered appropriate by policymakers, rather than used as a replacement for traditional models of corporate criminal liability entirely. The specific contexts in which such offences might be relevant is beyond the scope of this article. Instead, this article seeks to set out a model for a new type of corporate criminal offence for consideration by policymakers.

B *The ‘System of Conduct or Pattern of Behaviour’ Concept*

As noted above, the conceptual foundation adopted by the ALRC for system of conduct offences is the concept of a system of conduct or pattern of behaviour that exists in the civil regulatory provisions dealing with statutory unconscionability.¹¹⁵ The concept’s origins in statutory unconscionability may make it appear inapposite for broader application. However, the developing jurisprudence demonstrates that it can be used as a tool for assessing evidence of conduct — in particular, conduct arising from deficient corporate systems, procedures, processes, policies, patterns of work, and cultures — to determine whether the relevant conduct amounts to a system or pattern.

1 *Development*

The provisions in the *Australian Consumer Law* and *Australian Securities and Investments Commission Act 2001* (Cth) (*ASIC Act*) prohibiting unconscionable conduct reflect the system of conduct or pattern of behaviour concept.¹¹⁶ Provisions in both Acts state that ‘it is the intention of the Parliament that ... this section is capable of applying to a system of conduct or pattern of behaviour, whether or not a particular individual is identified as having been disadvantaged by the conduct or behaviour’.¹¹⁷ The express reference to a system of conduct or pattern of behaviour is one of several ‘interpretive principles’¹¹⁸ that were added to those statutes by the *Competition and Consumer Legislation Amendment Act 2011* (Cth).

¹¹⁴ Penny Crofts, Submission No 61 to ALRC, *Corporate Criminal Responsibility: Discussion Paper* 9.

¹¹⁵ *Australian Consumer Law* (n 7) s 21; *ASIC Act* (n 7) s 12CB. For a summary of the prohibition on statutory unconscionability and how it is established, see Michelle Sharpe, ‘“More Than a Feeling”: Finding Statutory Unconscionable Conduct’ 27(2) *Australian Journal of Competition and Consumer Law* 108. See also Jeannie Marie Paterson, Elise Bant and Matthew Clare, ‘Doctrine, Policy, Culture and Choice in Assessing Unconscionable Conduct under Statute: *ASIC v Kobelt*’ (2019) 13(1) *Journal of Equity* 81.

¹¹⁶ *Australian Consumer Law* (n 7) s 21; *ASIC Act* (n 7) s 12CB. The provision in the *Australian Consumer Law* is directed to unconscionable conduct in connection with goods and services, while that in the *ASIC Act* relates to financial services.

¹¹⁷ *Australian Consumer Law* (n 7) s 21(2)(b); *ASIC Act* (n 7) s 12CB(4)(b).

¹¹⁸ Explanatory Memorandum, *Competition and Consumer Legislation Amendment Bill 2010* (Cth) 19 [2.8].

In the context of statutory unconscionability, the use of the system of conduct or pattern of behaviour concept ‘ensures that the focus is on the conduct in question, as opposed to the characteristics of a particular person, or the effect of the impugned conduct on that person’.¹¹⁹ The introduction of this interpretive principle was designed to confirm a judicial interpretation that proof of statutory unconscionability did not require a specific victim to be identified.¹²⁰ Consequently, in its history and present usage, the concept is tied to the statutory norm of unconscionability.

The enactment of the ‘interpretive principle’ relating to systems of conduct or patterns of behaviour has led regulators to pursue what have been termed as ‘system’ cases — cases focused on how a corporation has run its business, rather than emphasising the impact on particular individuals.¹²¹ Under the existing provisions, system cases require the court to reach an evaluative judgment, on the evidence, about whether the impugned conduct can be characterised as a system of conduct or pattern of behaviour and whether that system or pattern can be characterised as unconscionable. An increasing number of these cases have been brought since 2011.¹²² Some have been successful.¹²³ Others have not.¹²⁴ Where proceedings have been unsuccessful, in some cases this has been due to the failure of the regulator to prove that the impugned conduct constituted a system of conduct or pattern of behaviour.¹²⁵ An example of this is *Unique International College Pty Ltd (in liq) v Australian Competition and Consumer Commission* (*‘Unique v*

¹¹⁹ Ibid 25 [2.24].

¹²⁰ *Australian Securities and Investments Commission v National Exchange Pty Ltd* (2005) 148 FCR 132, 140 [30], 140–1 [33], 143 [44] (Tamberlin, Finn and Conti JJ) (*‘National Exchange’*); *Australian Competition and Consumer Commission v EDirect Pty Ltd* (2012) 206 FCR 160, 184 [70], 185–6 [72]–[73] (Reeves J) (*‘EDirect’*).

¹²¹ Explanatory Memorandum (n 118) 25 [2.24]; *Kobelt v Australian Securities and Investments Commission* (2018) 352 ALR 689, 721 [181] (Besanko and Gilmour JJ).

¹²² Some cases were brought before the system of conduct or pattern of behaviour concept was expressly adopted by the Parliament: see, eg, *Australian Competition and Consumer Commission v IMB Group Pty Ltd* (2002) ATPR (Digest) ¶46–221; *National Exchange* (n 120); *Australian Competition and Consumer Commission v Keshow* (2005) ATPR (Digest) 46–265; *EDirect* (n 120). Note, however, that not all of these actions were successful.

¹²³ See, eg, *Australian Competition and Consumer Commission v ACN 117 372 915 Pty Ltd (in liq)* (2015) ATPR ¶42–498, affirmed on appeal: *NRM Corporation Pty Ltd v Australian Competition and Consumer Commission* (2016) ATPR ¶42–531; *Australian Competition and Consumer Commission v Harrison* [2016] FCA 1543; *Australian Competition and Consumer Commission v Lifestyle Photographers Pty Ltd* [2016] FCA 1538; *Australian Competition and Consumer Commission v Get Qualified Pty Ltd (in liq) (No 2)* (2017) ATPR ¶42–548; *Australian Competition and Consumer Commission v Ford Motor Company of Australia Ltd* (2018) 360 ALR 124; *Australian Competition and Consumer Commission v ABG Pages Pty Ltd* [2018] FCA 764; *Australian Competition and Consumer Commission v Geowash Pty Ltd (Subject to a Deed of Company Arrangement) (No 3)* (2019) 368 ALR 441; *Australian Competition and Consumer Commission v Cornerstone Investment Aust Pty Ltd (in liq) (No 4)* [2018] FCA 1408; *Australian Competition and Consumer Commission v Australian Institute of Professional Education Pty Ltd (in liq) (No 3)* [2019] FCA 1982 (*‘Institute of Professional Education (No 3)’*); *Australian Securities and Investments Commission v One Tech Media Ltd* (2020) 275 FCR 1; *Australian Securities and Investments Commission v AGM Markets Pty Ltd (in liq) (No 3)* (2020) 275 FCR 57 (*‘AGM Markets (No 3)’*); *Australian Competition and Consumer Commission v Quantum Housing Pty Ltd* (2021) 388 ALR 577.

¹²⁴ *Australian Competition and Consumer Commission v Woolworths Ltd* [2016] FCA 1472 (*‘ACCC v Woolworths’*); *Unique v ACCC* (n 10); *ASIC v Westpac (No 2)* (n 60); *Australian Securities and Investments Commission v Kobelt* (2019) 267 CLR 1 (*‘ASIC v Kobelt (HCA)’*).

¹²⁵ *Unique v ACCC* (n 10); *ASIC v Westpac (No 2)* (n 60).

ACCC’).¹²⁶ In that case, the regulator failed to prove that Unique’s enrolment processes amounted to a system or pattern that was unconscionable, establishing only that Unique’s conduct had been unconscionable in relation to the enrolment of six individual consumers. In other cases, the regulator has established that a system or pattern existed, but failed to establish that the system or pattern was unconscionable.¹²⁷ For example, in *Australian Securities and Investments Commission v Kobelt*, Mr Kobelt’s provision of book-up credit to customers of the general store was a system, but a majority of the High Court of Australia found that it was not unconscionable.¹²⁸ In some unsuccessful cases, it is harder to identify the reason for the failure, because analysis of the system or pattern concept and the statutory norm of unconscionability is closely intertwined.

Although the existing uses of the system of conduct or pattern of behaviour concept can apply to both individuals and corporations, in our view a broader application of the concept has particular resonance with corporate defendants due to the nature of contemporary corporate action. Given that the role of systems may be amplified in large corporations, it may be that a system of conduct offence is particularly apt for capturing the narrative of offending in the case of a larger corporate defendant where breaches of the law result from the internal workings of the company or their manifestations.

2 *Establishing a ‘System’ or ‘Pattern’*

What then is a ‘system of conduct’ or ‘pattern of behaviour’? And how can these be proved? The system of conduct or pattern of behaviour concept involves a characterisation of particular impugned conduct established by the evidence: should the conduct be assessed as comprising a ‘system’ or amounting to a ‘pattern’? That system or pattern is then assessed according to a statutory norm or standard to determine whether, by the system or pattern (that is, the combination of conduct), a contravention of the statutory provision has occurred.

As Paterson and Bant have argued, system cases ‘require a different case strategy’.¹²⁹ The focus of such a case ‘is not on the impact on a few individual consumers but on the way in which the business practice operates’.¹³⁰ Consequently, different evidentiary approaches are needed.

In *Unique v ACCC*, the Full Court of the Federal Court of Australia offered the following guidance regarding proof of a system or pattern:

A ‘system’ connotes an internal method of working, a ‘pattern’ connotes the external observation of events. These words should not be glossed. How a system or a pattern is to be proved in any given case will depend on the circumstances. It can, however, be said that if one wishes to move from the particular event to some general proposition of a system it may be necessary

¹²⁶ *Unique v ACCC* (n 10).

¹²⁷ *ACCC v Woolworths* (n 124); *ASIC v Kobelt (HCA)* (n 124).

¹²⁸ *ASIC v Kobelt (HCA)* (n 124).

¹²⁹ JM Paterson and E Bant, ‘Should Australia Introduce a Prohibition on Unfair Trading? Responding to Exploitative Business Systems in Person and Online’ (2021) 44(1) *Journal of Consumer Policy* 1, 8.

¹³⁰ *Ibid.*

for some conclusions to be drawn about the representative nature or character of the particular event.¹³¹

In that case, the Australian Competition and Consumer Commission ('ACCC') sought to establish a system case by reference to Unique's conduct in relation to six individual consumers. Although successful at first instance, the case was rejected on appeal — there was insufficient evidence to establish that the six consumers were representative of the class of 3600 potentially affected consumers and to conclude that there was a system or pattern.¹³² In reaching this conclusion, the Court indicated that it was not stating that to prove a system or pattern the regulator had to adduce evidence about a majority of consumers, or that the use of a representative sample to prove a system case was impermissible, or that individual cases could not be used to prove a system or pattern; all of these were possible methods of proof depending on the particular context.¹³³ What the Court emphasised was that whether a system or pattern exists is 'highly fact-specific, and will rely to a significant extent on the forensic exercise the regulator chooses to undertake to prove the existence of the system, as well as any forensic exercise the respondent undertakes by way of answer'.¹³⁴ In *Unique v ACCC*, the ACCC's system case — based on conduct toward particular individuals — required

evidence about either a material proportion of individual consumers; or evidence about how and why the individual consumers were chosen; or evidence about the representativeness of the individual consumers, or a combination of all three.¹³⁵

Without some evidence of this type, the ACCC could not establish a system from the six consumers.

Conversely, in *Australian Competition and Consumer Commission v Australian Institute of Professional Education Pty Ltd (in liq) (No 3)*,¹³⁶ Bromwich J held that the ACCC had established a system of conduct or pattern of behaviour that was unconscionable. The ACCC overcame the 'conceptual flaws'¹³⁷ in other cases such as *Unique v ACCC* because of a 'key conceptual difference': 'that the vectors of reasoning went in different directions'.¹³⁸ In 'fundamental contrast' to *Unique v ACCC*, the ACCC's case in *Institute of Professional Education (No 3)*:

was substantially based upon evidence directly going to AIPE's internal workings, as proven by AIPE's former employees, as well as business records such as enrolment records and data, enrolment forms and other documents, together with complaints and how they were handled. This evidence combined to give a reasonably pervasive sense of what was taking place, and its likely impact could thereby be ascertained on the balance of probabilities. Evidence from individual consumers was then used to demonstrate, by example, how this pattern or system played out at the enrolment coalface. The evidence of

¹³¹ *Unique v ACCC* (n 10) 654 [104]. This description of a 'system' or 'pattern' was referred to favourably by Nettle and Gordon JJ in *ASIC v Kobelt (HCA)* (n 124) 56 [143].

¹³² *Unique v ACCC* (n 10) 655 [110]–[111].

¹³³ *Ibid* 666 [151]–[152].

¹³⁴ *Ibid* 666 [150].

¹³⁵ *Ibid* 670 [162].

¹³⁶ *Institute of Professional Education (No 3)* (n 123).

¹³⁷ Paterson and Bant (n 129) 7.

¹³⁸ *Institute of Professional Education (No 3)* (n 123) 66 [163] (Bromwich J).

the individual consumer witnesses was thus helpful and made for a stronger case for the applicants, but was not indispensable and not used as evidence that of itself was representative of the system or pattern.¹³⁹

Therefore, although *Unique v ACCC* in no way forecloses proof of a system case through individual cases, *Institute of Professional Education (No 3)* shows another method of proving a system or pattern: through primary reliance on internal evidence, including, in that case, the evidence of ‘three former relatively senior employees’.¹⁴⁰

More recently, Beach J in *Australian Securities and Investments Commission v AGM Markets Pty Ltd (in liq) (No 3)* noted that it was ‘important not to confuse ... the concepts of “system of conduct” and “pattern of behaviour”, and ... the mode of proof of such concepts’.¹⁴¹ His Honour explained that a ‘pattern of conduct’ related to the ‘external manifestation of behaviour’.¹⁴² As to proof of a pattern, Beach J said that the number of instances that must exist before a pattern is evidenced ‘depends upon the context’ and

Clearly, two or more instances of identical or similar behaviour may be sufficient to infer and discern a pattern. ...

Numerous like instances with no counter-examples would clearly be sufficient to display a pattern. ...

A pattern may still exist, notwithstanding the [existence of] exceptions. ...

[T]he greater the individual differences of the instances, being differences that have real significance to the characterisation of unconscionability, the greater the number of instances that may be required to justify the extrapolation ...

Some parts of [a respondent’s] conduct may manifest a pattern, other parts not. A ‘pattern of behaviour’ may be sufficiently found in relation to a part.¹⁴³

As to a system of conduct, Beach J noted that a system was not simply a similar pattern of behaviour.¹⁴⁴ His Honour offered the following guidance:

‘system’ connotes something designed or intended in its structure ...

‘system’ is usually saying something about the internal structure, for example, internal working, of whatever it is that has produced or reflects the conduct.

...

The gist is organisation and connection. ...

‘system of conduct’ mean[s] that each element of individual conduct is directly connected in a structured and intended way one to the other [and] ‘system of conduct’ focus[es] on the system as being the underlying internal structure or method of procedure, organisation or administration of the respondent ...¹⁴⁵

The guidance provided by these decisions, together with that in the other decided cases, amounts to an increasing body of jurisprudence on establishing a

¹³⁹ Ibid.

¹⁴⁰ Ibid 66 [164].

¹⁴¹ *AGM Markets (No 3)* (n 123) 122 [385].

¹⁴² Ibid 122 [386].

¹⁴³ Ibid 122 [387]–[388].

¹⁴⁴ Ibid 122 [388].

¹⁴⁵ Ibid 123 [389]–[391].

system of conduct or pattern of behaviour. At the same time, as the Full Federal Court said in *Unique v ACCC*, the ‘words should not be glossed’ — ‘[h]ow a system or a pattern is to be proved in any given case will depend on the circumstances’.¹⁴⁶ The ultimate question is whether the evidence establishes conduct that can properly be characterised as a system of conduct or pattern of behaviour. Importantly, we consider that this technique of assessing and characterising conduct as a system or pattern has potential utility in the characterisation of contemporary corporate action more generally, including for the purposes of a system of conduct offence, as proposed by the ALRC.

C *A Broader Role for the Concept*

Although the system of conduct or pattern of behaviour concept is presently linked to statutory unconscionability, the case law on the concept illuminates how it might operate as a tool for characterising corporate misconduct more broadly, such that it has potential utility in corporate regulation generally.¹⁴⁷ If anything, the concept may be easier to apply to misconduct outside of statutory unconscionability.¹⁴⁸

Specifically, the concept, as a tool for characterising a particular concatenation of conduct in context, has potential utility as the foundation for a novel type of offence directed at systematic misconduct, where the focus should be on the nature of the corporate conduct and its systematic characteristics. Existing offences do not adequately capture this narrative of offending — the criminality of systematic misconduct. At the same time, the capacity of contemporary corporations to act through systems, processes, policies, procedures, and culture, is significant, as discussed above.

More fundamentally, and as noted above, Paterson and Bant have argued that the system of conduct or pattern of behaviour concept represents a different type of approach to regulation.¹⁴⁹ Paterson, Bant, and Clare have also opined, with reference to the Royal Commission into Misconduct in the Banking, Superannuation and Financial Services Industry, that the system of conduct or pattern of behaviour concept ‘has the potential to provide a powerful regulatory weapon against ... sophisticated corporate wrongdoers’.¹⁵⁰ They further observed that:

Given the significant difficulties in attributing dishonest or predatory intentions in the corporate context, the statutory provisions relating to systems provide a potentially important route to holding large and complex corporate wrongdoers responsible for unconscionable conduct.¹⁵¹

¹⁴⁶ *Unique v ACCC* (n 10) 654 [104].

¹⁴⁷ This broader use of the concept, beyond statutory unconscionability, may also remove some of the difficulties in its use, as the underlying norm that the system or pattern is alleged to contravene may be simpler in its content and application.

¹⁴⁸ As illustrated by the practical examples in *ALRC Corporate Criminal Responsibility Report* (n 3) 283–93 [7.46]–[7.52].

¹⁴⁹ Paterson and Bant (n 129) 15–16.

¹⁵⁰ Paterson, Bant and Clare (n 115) 104.

¹⁵¹ *Ibid.*

As the remarks of Paterson, Bant and Clare make clear, given the characteristics of contemporary corporate misconduct and the difficulties of traditional approaches to corporate criminal responsibility based on principles of attribution, there is potential for use of the concept of system of conduct or pattern of behaviour as the foundation for a novel type of offence tailored to a corporation. The ALRC has already adopted the concept as the foundation for such a type of offence.

D *The ALRC's Draft Model Offence for Systematic Misconduct*

The ALRC's eighth recommendation in its *Corporate Criminal Responsibility Report* was that 'the Australian Government should introduce offences that criminalise contraventions of prescribed civil penalty provisions that constitute a system of conduct or pattern of behaviour by a corporation'.¹⁵²

The ALRC explained that the recommended offence would provide a means of capturing 'systematic misconduct that would otherwise have to be dealt with through civil enforcement'.¹⁵³ The ALRC recommended the enactment of this system of conduct offence in specific regulatory contexts as an adjunct to ordinary criminal offences (such offences being applied to corporations through methods of attribution).¹⁵⁴ The ALRC did not recommend replacing responsibility based on attribution for ordinary criminal offences.¹⁵⁵ Consistent with its analysis that existing offences tailored to corporations, such as duty-based offences and failure to prevent offences, have been more effective in practice in addressing corporate misconduct, the ALRC described the recommended offence as an additional option for legislators. It would be used where there is a policy objective that involves addressing systematic corporate misconduct.¹⁵⁶

The offence contemplated by the ALRC has the following features:

- the use of a 'system of conduct or pattern of behaviour' concept;
- the requirement that at least two ... contraventions [of prescribed civil penalty provisions] have resulted; and
- the need to prove the particular fault elements.¹⁵⁷

Part of the ALRC's rationale for the recommendation was that 'there is a need to effectively design regulatory provisions that address contravening business systems and practices'.¹⁵⁸ Such provisions recognise that corporations act collectively. The recommended offence formed part of a suite of recommendations that, taken together, would see a decriminalisation of many of the low-level offences currently applicable to corporate misconduct, with a renewed focus on civil penalties. In the

¹⁵² *ALRC Corporate Criminal Responsibility Report* (n 3) 15 recommendation 8.

¹⁵³ *Ibid* 271 [7.9].

¹⁵⁴ *Ibid* 283 [7.46]–[7.47].

¹⁵⁵ *Ibid*.

¹⁵⁶ *Ibid*.

¹⁵⁷ *Ibid* 270 [7.7].

¹⁵⁸ *Ibid* 271 [7.9], citing Jeannie Marie Paterson and Elise Bant, 'Unfair, Unjust and Unconscionable Conduct in Consumer Contracting: Designing Effective Law Reform in Australia' (Conference Paper, Melbourne Law School Obligations Group Annual Conference, 5–6 December 2019).

context of this package of recommendations, the system of conduct offence would address the ALRC's concern that civil penalties may be seen as a 'cost of doing business'.¹⁵⁹ Historically, the response to this risk has been to ratchet up the applicable penalty.¹⁶⁰ A large civil penalty may not be sufficient to reflect the moral opprobrium that should attach to systematic contraventions of civil prohibitions, and, given the unique expressive power of the criminal law, it was considered appropriate that such breaches of the law could result in potential criminal liability.

V Framing System of Conduct Offences to Enhance Corporate Criminal Law

The final part of this article considers how the system of conduct or pattern of behaviour concept might be utilised in contexts beyond the draft model offence recommended by the ALRC: first, as the foundation of a broader model of system of conduct offence tailored to the reality of corporate action; and, second, as an evidentiary framework of analysis.

As noted above, the ALRC's recommended offence would apply to systems of conduct or patterns of behaviour engaged in by corporations that result in breaches of two or more civil penalty provisions. The context for this recommendation was that the ALRC was recommending a significant winding back of the criminal law as it applies to corporations, recognising the pre-eminence of civil penalties as a tool of corporate regulation in Australia. As we have noted, the recommended offence was designed to ensure that corporations did not treat civil penalty provisions as merely a cost of doing business and to enable an escalation of enforcement responses where systematic misconduct was observed, as evidenced by multiple civil penalty breaches.¹⁶¹

There are arguably two further potential uses of the system of conduct or pattern of behaviour concept beyond that recommended by the ALRC.¹⁶²

A *Uncoupling the Offence from Civil Penalty Contraventions*

The first further application involves uncoupling the recommended offence from the need to prove multiple breaches of civil penalty provisions. In this way, the system of conduct offence would sit alongside the civil penalty provision in a form of 'dual-track' regulation. As the ALRC noted, dual-track regulation is common in many

¹⁵⁹ ALRC Corporate Criminal Responsibility Report (n 3) 271 [7.11].

¹⁶⁰ See, eg, *Australian Competition and Consumer Commission v Yazaki Corporation* (2018) 262 FCR 243, 300 [259] (Allsop CJ, Middleton and Robertson JJ); *Singtel Optus Pty Ltd v Australian Competition and Consumer Commission* (2012) 287 ALR 249, 266 [68] (Keane CJ, Finn and Gilmour JJ). It was also a justification for the increase in maximum penalties in relation to a number of civil penalty provisions as a consequence of the ASIC Enforcement Review: Explanatory Memorandum, Treasury Laws Amendment (Strengthening Corporate and Financial Sector Penalties) Bill 2018 (Cth) 22 [1.32]–[1.33], 23–4 [1.40]–[1.41], 39 [1.108], 47–8 [1.144], 53 [1.154], 64 [1.214].

¹⁶¹ ALRC Corporate Criminal Responsibility Report (n 3) 271 [7.11].

¹⁶² The ALRC referred to such a possibility in its Final Report, but did not make any recommendations: *ibid* 283 [7.45].

areas of Commonwealth regulation, where particular conduct is subject to both a civil penalty provision and a criminal offence.¹⁶³ Currently, where there is any distinction between the civil penalty provision and the criminal offence, it is that the criminal offence has additional fault elements such as intention and recklessness.¹⁶⁴ As outlined above, what constitutes criminal intent on the part of a corporation is particularly problematic in terms of securing accountability, arguably more so where the associated civil wrong has no fault element and a higher penalty.

Dual-track regulation provides the regulator with flexibility and ‘in accordance with responsive regulation theory, it enables selection by the regulator of the pathway considered most appropriate for achieving compliance by a particular corporation or industry’.¹⁶⁵ Thus, the regulator could pursue a corporation under the criminal law if there was evidence of a system of conduct or pattern of behaviour leading to the corporate misconduct. It would be the evidence of systems or patterns within the corporation’s operations that would justify the imposition of a criminal sanction as opposed to a civil penalty, which would be more appropriate in circumstances of more ad hoc non-compliance.

Under the approach we propose, charging fees for no service in the financial services industry could be a criminal offence if undertaken by a corporation in a manner that can be characterised as amounting to a system of conduct or pattern of behaviour. Similarly, the underpayment of wages could amount to a criminal offence if there is evidence of systems of conduct or patterns of behaviour that demonstrate that such underpayments are not simply the result of isolated errors. This example connects with recent debates about the criminalisation of wage theft as an effective regulatory response.¹⁶⁶ It could be argued that it is the systematic nature of the wrongdoing that makes the conduct sufficiently wrongful so as to warrant criminalisation.

Interestingly, in Queensland wage theft was recently criminalised through amendments to the definitions of the offences of stealing and fraud under the *Criminal Code Act 1899* (Qld) (*‘Criminal Code (Qld)’*).¹⁶⁷ Arguably, this does not acknowledge that, in reality, the underpayment (or non-payment of wages) may be due to a range of circumstances including clerical error(s), pay system design and payment of wages below the minimum legal entitlements. Moreover, it makes the criminal offence of wage theft difficult to prove, particularly in a corporate context, as it is necessary to prove a fraudulent intention. Given the range of explanations for

¹⁶³ Ibid 87–8 [3.48]–[3.49], 172–6 [5.17]–[5.26].

¹⁶⁴ Ibid 175–6 [5.24].

¹⁶⁵ Ibid 176 [5.25]. Examples include *Corporations Act* (n 71) ss 601ED(5), 670A, 728, 792B, 853F, 904C, 905A, 911A, 911B, 920C, 922M, 952E, 952H, 993D, 1020A, 1021E, 1021G, 1309; *National Consumer Credit Protection Act 2009* (Cth) sch 1, ss 24, 154, 155–156, 174, 179U, 179V; *Insurance Contracts Act 1984* (Cth) s 33.

¹⁶⁶ See, eg, Melissa Kennedy, Submission to Attorney-General’s Department (Cth), *Discussion Paper on Improving Protections of Employees’ Wages and Entitlements: Strengthening Penalties for Non-Compliance* (2019).

¹⁶⁷ Criminal Code and Other Legislation (Wage Theft) Amendment Bill 2020 amended s 391 of the *Criminal Code* (Qld) (n 2) with respect to stealing and s 408C with respect to fraud. For analysis, see Joachim Dietrich and Matthew Raj, ‘Criminalising “Wage Theft” in Australia: Property, Stealing, and Other Concepts’ (2021) 45(4) *Criminal Law Journal* 218.

incorrect wages, coupled with the attribution issues highlighted in this article, the Queensland offence may be little more than a symbolic enactment.

Kennedy has argued that criminal sanctions for wage theft ‘should be reserved for the most serious cases ... and be used as part of a multi-faceted regulatory response’.¹⁶⁸ In our view, this is where a system of conduct offence could play a role in addressing wage theft by corporate employers. Arguably, what legitimately turns the civil wrong of underpaying wages into criminal conduct is the systematic nature of the misconduct — the development of a business model that routinely underpays employees. If that conception of criminality is accepted, then the system of conduct offence could be an apt solution and there is no need to link that offence to the breach of a civil penalty provision or look to a corporate state of mind. It is the nature of the misconduct that renders it deserving of the expressive force of the criminal law, over and above a finding of a civil penalty contravention.

This is the approach that was taken in the wage theft offence that was proposed in the Fair Work (Supporting Australia’s Jobs and Economic Recovery) Bill 2020 (Cth), which provided that ‘[a]n employer commits an offence if the employer dishonestly engages in a *systematic pattern* of underpaying one or more employees’.¹⁶⁹ In contrast to the approach proposed by the ALRC and this article, however, the Bill provided a series of matters to which a court may have regard in determining whether the employer engaged in a systematic pattern of underpaying employees. These included

the number of underpayments; the period over which the underpayments occurred; the number of employees affected by the underpayments; the employer’s response, or failure to respond, to any complaints made about the underpayments; whether the employer failed to comply with a requirement ... in relation to an employee record relating to the underpayments; and whether the employer failed to comply with a requirement ... in relation to a pay slip relating to the underpayments.¹⁷⁰

The proposed offence did not use the system of conduct or pattern of behaviour concept discussed above, but it did demonstrate a recognition by policymakers that systematic misconduct may be considered capable of shifting civil contraventions into the criminal sphere.

On another note, where achieving compliance is sufficiently important so as to justify the existence of strict or absolute liability offences, rather than merely civil penalty provisions, there would also be scope for the legislature to enact an applicable system of conduct offence to cover the proscribed conduct, if policymakers took the view that the systematicity of misconduct was also of particular concern in the relevant context. This would enable an effective response to misconduct that is aggravated due to its systematic nature.

¹⁶⁸ Kennedy (n 166) 8.

¹⁶⁹ Fair Work (Supporting Australia’s Jobs and Economic Recovery) Bill 2020 (Cth) sch 5 cl 46, proposing to introduce s 324B of the *Fair Work Act 2009* (Cth) (emphasis added). Note that, ultimately, the proposed offence was omitted from the legislation that was actually enacted by the Australian Parliament.

¹⁷⁰ Ibid, proposed s 324B(5).

In our view, and consistently with the ALRC's recommendation, any broader use of the system of conduct model should only be applied in specific contexts, where concern about the potential for systematic misconduct means that it is an appropriate regulatory response. However, if this option was taken to its extreme and system of conduct offences were enacted in a wide range of contexts, together with other offences tailored to corporations and also strict liability offences, then they could effectively replace traditional approaches based on attribution, as was proposed to the ALRC by Bant:

given the difficulties around corporate attribution, one direction for future reform might be to take seriously the original position taken by the courts, which was that as artificial entities, corporations lacked 'minds,' and instead focus on the objective quality of the conduct of corporations.¹⁷¹

Under the model put forward by Bant, rather than proving fault elements through principles of attribution, it would be necessary to instead examine the corporation's conduct objectively. This would require an examination of what happened routinely within the corporation, rather than the aspirations described in mission statements and compliance manuals.

The ALRC did not go this far in the *Corporate Criminal Responsibility Report*. Looking at the enforcement data together with the theoretical concepts relating to corporate fault, the ALRC's central argument was that improved principles of attribution should be augmented by specific offences that go to the nature of corporate misconduct in particular instances and industries.¹⁷² These offences would not negate entirely the need for a model of corporate attribution. The ALRC's recommended system of conduct or pattern of behaviour model offence would not replace attribution principles and, in addition, it would still require proof of fault elements through traditional methods.¹⁷³ Furthermore, corporate misconduct sometimes involves one-off instances — particularly with regard to fraud or dishonesty offences — that would not reach the threshold of systematic misconduct. What a system of conduct offence may do, like all offences tailored to corporations, is provide regulators and prosecutors with another tool for addressing corporate misconduct and for appropriately capturing the particular nature of offending. The evidence obtained by the ALRC as to the frequency of corporate prosecutions in relation to duty-based offences, another type of specific offence discussed earlier in this article, suggests that this tailoring of offences is more effective than traditional approaches to corporate criminal liability based on attribution.¹⁷⁴

That tailoring is urgently needed. With just 13 prosecutions over 10 years under the *Criminal Code* (Cth),¹⁷⁵ the lack of prosecutions is an indictment of a

¹⁷¹ Bant (n 68) 6.

¹⁷² *ALRC Corporate Criminal Responsibility Report* (n 3) 283 [7.46]–[7.47].

¹⁷³ *Ibid* 272–3 [7.12]–[7.13]. As to the possibility of reforming approaches to conceptualising corporate states of mind, Bant has recently posited a new approach to corporate culpability in the form of “‘systems intentionality’”, in which the corporate mind is manifested through its corporate systems, policies and patterns of behaviour’: Bant (n 109) 354–5. Such a ‘model would stand alongside and supplement existing general law and statutory attribution rules’: at 355. See also Bant and Paterson (n 109).

¹⁷⁴ *ALRC Corporate Criminal Responsibility Report* (n 3) 218–19 [6.5].

¹⁷⁵ *Ibid* 105 [3.91].

collective failure to address corporate misconduct. While offences tailored to corporations, including the system of conduct offence recommended by the ALRC, can be an important part of the mix, they are not a panacea for other limitations relevant to holding errant corporate entities responsible — notably, our existing methods of attribution, limited sentencing options, and investigative limitations. The ALRC's *Corporate Criminal Responsibility Report* makes recommendations for improving these parts of the regulatory mix as well.

B *Use as an Evidentiary Framework of Analysis*

The second wider application of the system of conduct or pattern of behaviour concept beyond that recommended by the ALRC, is as an evidentiary tool that could improve methods of corporate criminal liability based on attribution, through providing guidance as to how to establish authorisation or permission through corporate culture. Much of the criticism of pt 2.5 of the *Criminal Code* (Cth), and its corporate culture provisions in particular,¹⁷⁶ is that it is not clear how those provisions should be interpreted or how a corporate culture may, by analogy, be said to authorise or permit the commission of the physical elements of the offence.¹⁷⁷ Dixon has argued that the corporate culture provisions involve 'evidentiary burdens too high to meet with any practical certainty',¹⁷⁸ and Justice French described the provisions as 'extraordinarily wide and vague'.¹⁷⁹ There is also a paucity of practical guidance as to how to prove a case reliant on corporate culture. The Commonwealth Director of Public Prosecutions has noted the 'little available judicial authority' and observed that the provisions are 'largely untested', although there are active cases relying on the corporate culture provisions.¹⁸⁰ We now have, as outlined above, a significant and growing body of jurisprudence as to what can constitute a system of conduct or pattern of behaviour and what evidence is necessary to establish that they exist. What is the culture of a corporation and how is it manifested? Arguably, it is manifested in the ways of working, the patterns of behaviour, and systems of conduct of the corporation. Thus, the burgeoning jurisprudence on systems of conduct and patterns of behaviour could be used to provide guidance as to how to establish a corporate culture for the purposes of pt 2.5 of the *Criminal Code* (Cth). There is significant support for the corporate culture provisions and the model of corporate fault that they embody.¹⁸¹ The system or pattern concept could provide an evidentiary framework of analysis for operationalising the more nebulous concept

¹⁷⁶ *Criminal Code* (Cth) (n 1) ss 12.3(2)(d), 12.3(6).

¹⁷⁷ Corporate culture is a mechanism for proving this authorisation or permission: *ALRC Corporate Criminal Responsibility Report* (n 3) 59–60 [2.45]–[2.46].

¹⁷⁸ Olivia Dixon, 'Corporate Criminal Liability: The Influence of Corporate Culture' (Sydney Law School Legal Studies Research Paper No 17/14, The University of Sydney, February 2017) 16.

¹⁷⁹ The Hon Justice RS French, 'The Culture of Compliance — A Judicial Perspective' (Conference Paper, Australian Compliance Institute National Conference, 3–5 September 2003).

¹⁸⁰ *ALRC Corporate Criminal Responsibility Report* (n 3) 247–8 [6.115], quoting the Commonwealth Director of Public Prosecutions ('CDPP'), Submission No 56 to ALRC, *Corporate Criminal Responsibility: Discussion Paper* (7 February 2020). The ALRC was made aware of one finalised case involving the corporate culture provisions: *R v Potter* (2015) 25 Tas R 213.

¹⁸¹ *ALRC Corporate Criminal Responsibility Report* (n 3) 246–50 [6.109]–[6.120], citing the views of Bant, Crofts, and Clough and Mulhern, together with Allens, the CDPP, and ASIC.

of corporate culture in the concrete circumstances of an appropriate individual prosecution.¹⁸²

VI Conclusion

Thirty years ago, Fisse described the ‘attribution of criminal liability to corporations’ as ‘one of the blackest holes in criminal law’.¹⁸³ As Parts I and II of this article sought to demonstrate, much of the problem with traditional approaches to corporate criminal liability arises from the historical evolution of corporate criminal liability, and the inherent difficulty of adapting the criminal law for application to corporations. One avenue that policymakers have increasingly taken in order to overcome the limitations of traditional approaches and effectively address corporate misconduct has been to develop offences that better reflect the nature of corporate action. Building on the work of the ALRC’s *Corporate Criminal Responsibility Report*, this article has argued that the enactment of a novel type of offence tailored to corporations, and targeting corporate systems of conduct and patterns of behaviour, would enhance the law’s ability to respond to contemporary corporate misconduct. Much misconduct is often systematic in nature due to the complexity of modern corporations. Enactment of system of conduct offences would enhance corporate accountability, and also level the playing field for all corporations, by ensuring that those that adopt systems and processes in breach of the law do not derive an unjust advantage over those that comply with the law.

¹⁸² See also the model of ‘systems intentionality’ developed in Bant (n 109) 381–8.

¹⁸³ Brent Fisse, ‘The Attribution of Criminal Liability to Corporations: A Statutory Model’ (1991) 13(3) *Sydney Law Review* 277, 277.

The “Constitutional” Value of the Racial Discrimination Act 1975 (Cth)

Alice Taylor*

Abstract

Since its passage, the *Racial Discrimination Act 1975* (Cth) (*RDA*) has been considered ‘special’. This is despite the fact that it is a piece of ordinary legislation, capable of being amended or limited by the legislature. This article considers the nature of this ‘specialness’. I assess whether the *RDA* can be classified as a ‘constitutional’ statute in the Australian context. I argue that, utilising a range of definitions, the *RDA* can be classed as a piece of ‘constitutional’ legislation, but that this status has no discernible effect on producing effective and substantive protection from discrimination on the basis of race.

Please cite this article as:

Alice Taylor, ‘The “Constitutional” Value of the *Racial Discrimination Act 1975* (Cth)’ (2021) 43(4) *Sydney Law Review* 519.



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC 4.0).

As an open access journal, unmodified content is free to use with proper attribution. Please email sydneylawreview@sydney.edu.au for permission and/or queries.

© 2021 Sydney Law Review and author. ISSN: 1444–9528

* Assistant Professor, Bond University, Robina, Queensland, Australia. Email: altaylor@bond.edu.au; ORCID iD: <https://orcid.org/0000-0002-9663-7676>. I thank the reviewers for their helpful feedback, all errors are my own.

I Introduction

The *Racial Discrimination Act 1975* (Cth) ('*RDA*') occupies a special place in Australia's non-discrimination regime. When the *RDA* was proclaimed in 1975, Prime Minister Gough Whitlam commented that while the Bill did not have the same 'rhetorical grandeur' of the rights-protecting documents of the United States, he hoped that it would have the same compelling and lasting force.¹ Whitlam could make this statement because of the *RDA*'s potential as a rights-protecting document. In 1982, Sir Harry Gibbs made this emphasis on rights explicit when he concluded that s 10 of the *RDA* implemented 'a bill of rights ... which is effective[ly] entrenched against the States',² thereby asserting the Act's almost constitutional status.

The assertion of the 'special' place of the *RDA* in Australia's legislative landscape is also present in academic scholarship. Williams and Reynolds suggest that the *RDA* has 'political importance attached to it over and above almost any other piece of Commonwealth legislation'.³ Given the *RDA*'s political significance, in the context of Indigenous constitutional recognition, Lino has suggested that rather than pursuing constitutional change to incorporate a non-discrimination clause in the *Australian Constitution*, a manner and form provision could be embedded in the *RDA* to give the Act stronger constitutional force.⁴ Though acknowledging its limitations, the *RDA* has been cited as an example of Australian 'values' in a consideration of global constitutional values by Saunders and Donaldson,⁵ and an example of an Australian 'constitutional' statute by Stephenson.⁶

However, notwithstanding the 'special' nature of the *RDA*, little work has been conducted on whether its political significance has had any discernible effect on the interpretation and effectiveness of the *RDA* in eliminating racially discriminatory conduct and laws. Thus, in this article, I interrogate the 'special' or even 'constitutional' nature and force of the *RDA*. This is to determine two interrelated issues. First, this article will determine whether the *RDA* can be classified as a 'special' or even 'constitutional' form of legislation within the Australian context. Second, it will consider the effect of this possible classification on the interpretation of unlawful racial discrimination. The effect on interpretation is important because in the context of 'constitutional' or 'quasi-constitutional' statutes elsewhere, notably in Canada, the effect of declaring human rights

¹ Gough Whitlam, 'Proclamation of *Racial Discrimination Act*: Speech by the Prime Minister at a Ceremony Proclaiming the *Racial Discrimination Act 1975*' (Speech, Canberra, 31 October 1975) [5].

² Sir Harry Gibbs, 'Eleventh Wilfred Fullagar Memorial Lecture: The Constitutional Protection of Human Rights' (1982) 9(1) *Monash University Law Review* 1, 13.

³ George Williams and Daniel Reynolds, 'The *Racial Discrimination Act* and Inconsistency under the *Australian Constitution*' (2015) 36(1) *Adelaide Law Review* 241, 242.

⁴ Dylan Lino, 'Thinking outside the *Constitution* on Indigenous Constitutional Recognition: Entrenching the *Racial Discrimination Act*' (2017) 91(5) *Australian Law Journal* 381, 384 ('Thinking outside the *Constitution*').

⁵ Cheryl Saunders and Megan Donaldson, 'Values in Australian Constitutionalism' in Dennis Davis, Alan Richter and Cheryl Saunders (eds), *An Inquiry into the Existence of Global Values: Through the Lens of Comparative Constitutional Law* (Hart Publishing, 2015) 15, 36.

⁶ Scott Stephenson, 'The Rise and Recognition of Constitutional Statutes' in Richard Albert and Joel Colón-Ríos (eds), *Quasi-Constitutionality and Constitutional Statutes: Forms, Functions and Applications* (Routledge, 2019) 35, n 33.

legislation to be quasi-constitutional has been significant in pursuing a ‘broad, beneficial and purposive’ interpretation of the legislation.⁷ Without such effect on interpretation, the utility of designating the *RDA* as ‘constitutional’ is questionable.

In this article, I utilise the case law on s 10 of the *RDA* to consider the ‘constitutional’ value of the *RDA*. Section 10 of the *RDA* prohibits discriminatory laws on the basis of race and it is outlined in the following terms:

If, by reason of, or of a provision of, a law of the Commonwealth or of a State or Territory, persons of a particular race, colour or national or ethnic origin do not enjoy a right that is enjoyed by persons of another race, colour or national or ethnic origin, or enjoy a right to a more limited extent than persons of another race, colour or national or ethnic origin, then, notwithstanding anything in that law, persons of the first-mentioned race, colour or national or ethnic origin shall, by force of this section, enjoy that right to the same extent as persons of that other race, colour or national or ethnic origin.⁸

I consider s 10 for this article because it is a unique provision in the discrimination law landscape. There are no similar provisions in any other Commonwealth discrimination statutes.⁹ These statutes are focused on discriminatory conduct rather than laws. And in an international context, provisions like s 10 are generally found in constitutional frameworks rather than ordinary statutes. Consequently, my approach in this article — focusing on one specific aspect of s 10 of the *RDA* — is a narrow one. I have adopted this focus both for practical and conceptual reasons. From a practical perspective, the focus on s 10 allows for a clear, detailed analysis in this article built on a foundation of High Court of Australia jurisprudence. From a conceptual basis, the distinctive place of s 10 within the Australian discrimination law context, with its focus on distinctions made in law rather than discriminatory conduct, provides a solid foundation to assess the ‘constitutionality’ of the *RDA* given its clear commonality with statutory and constitutional human rights instruments in other jurisdictions.

I argue, utilising a number of prominent definitions, that the *RDA* can be classed as a form of ‘constitutional’ legislation due to the combination of its subject matter and effect on state legislation-making powers. However, unlike in other jurisdictions such as Canada, I argue such a classification has achieved little due to the lack of purposive interpretation of the rights contained in the *RDA*. As a consequence, few pieces of legislation have been found to be inconsistent.

I present this argument in three parts. First, in Part II, I examine both the aspirational and practical potential of the *RDA* to demonstrate that it has a ‘special’ place in Australian law. The *RDA*’s history and passage and its treatment by succeeding legislatures demonstrates this. Practically, the Act has a powerful effect on state as well as Commonwealth legislatures in terms of regulating legislative and executive action. In this sense, the *RDA* can be considered a ‘special’ form of legislation.

⁷ John Helis, *Quasi-Constitutional Laws of Canada* (Irwin Law, 2018) 3.

⁸ *Racial Discrimination Act 1975* (Cth) s 10(1) (*‘RDA’*).

⁹ There is no equivalent provision in the *Sex Discrimination Act 1985* (Cth), the *Disability Discrimination Act 1992* (Cth) or the *Age Discrimination Act 2004* (Cth).

However, despite the recognition of the *RDA* as a ‘special’ piece of legislation more broadly, the interpretation by the judiciary has been a disappointment. Thus, in Part III I consider the case law from the High Court and appellate court decisions from Queensland and the Northern Territory to demonstrate that this special status has had no noticeable effect on the interpretation of the *RDA*, focusing specifically on the interpretation of s 10 of the *RDA* that guarantees equality before the law. In particular, I argue that the interpretation of the terms of s 10 and specifically what is considered a ‘distinction on the basis of race’, leads to an interpretation of the *RDA* that is disconnected from the lived reality of race-based disadvantage that the *RDA* was designed to combat.¹⁰ In Part IV, I contextualise the approach adopted with respect to s 10 within the broader literature on constitutional culture and the use of values in interpretation, particularly the value of equality to demonstrate that though the interpretation of s 10 is unusual when compared to its foreign counterparts, it is consistent with the interpretation of values in a constitutional setting in Australia.

II Defining the *Racial Discrimination Act 1975* (Cth) as a ‘Constitutional’ Statute

Though the concept of ‘constitutional’ statutes has gained little traction in Australian High Court jurisprudence,¹¹ the persistent references to the *RDA*’s importance and ‘special’ place in Australian law should give pause to consider whether, through its function or subject matter, the *RDA* has a status above an ‘ordinary’ statute. In this section, I chart the perceived ‘specialness’ of the *RDA* against some of the various definitions of ‘constitutional’ or ‘quasi-constitutional’ legislation to demonstrate that the *RDA* can be considered a form of ‘constitutional’ legislation, both in its form and in its effect.

The *RDA* has been presented by politicians and judges extra-curially, as a ‘special’ form of legislation.¹² The description of the *RDA* as special has been justified on two bases. The first is on the basis of the subject matter, in that discrimination legislation is reflective of fundamental values.¹³ The second is due to the fact that, unlike other Commonwealth discrimination legislation, the *RDA* confines lawmaking powers, particularly of state legislatures due to s 10.¹⁴

In other jurisdictions, such as the United Kingdom (‘UK’) and Canada, it is more common for judges to declare legislation as ‘constitutional’ (in the UK),¹⁵ or

¹⁰ Beth Gaze, ‘Has the *Racial Discrimination Act* Contributed to Eliminating Racial Discrimination? Analysing the Litigation Track Record 2000–04’ (2005) 11(1) *Australian Journal of Human Rights* 171, 171–2 (‘*Racial Discrimination Act*’).

¹¹ *Cadia Holdings Pty Ltd v New South Wales* (2010) 242 CLR 195, 217–18 (French CJ) (‘*Cadia*’).

¹² See Whitlam (n 1) [5]; Gibbs (n 2) 13.

¹³ Saunders and Donaldson (n 5) 36; Simon Evans, ‘Why is the *Constitution* Binding? Authority, Obligation and the Role of the People’ (2004) 25(1) *Adelaide Law Review* 103, 116–17.

¹⁴ Williams and Reynolds (n 3) 242.

¹⁵ Farrah Ahmed and Adam Perry, ‘Constitutional Statutes’ (2017) 37(2) *Oxford Journal of Legal Studies* 461, 464.

as ‘quasi-constitutional’ (in Canada).¹⁶ In Canada, in particular, there is now a list of statutes that the courts have designated as quasi-constitutional.¹⁷ This list includes the *Canadian Bill of Rights*,¹⁸ the federal and provincial human rights statutes (discrimination laws), privacy legislation, freedom of information legislation and statutes relating to language rights.¹⁹ For the purpose of this article, I start from the position that the terms ‘quasi-constitutional’ and ‘constitutional’ used in the scholarship and the case law are ostensibly referring to the same kinds of legislation. Constitutional or quasi-constitutional legislation is legislation that sits at a level between a ‘capital-C style’ Constitution and an ordinary statute.²⁰ Nevertheless, judges in both the UK and Canada have often failed to provide a definition or outline of what features give a statute this particular designation.²¹ The case law and the literature provide three possible definitions of a constitutional statute, which will be considered here: first, statutes where the subject matter concerns ‘fundamental’ rights;²² second, statutes that, due to both their subject matter as well as the political circumstances surrounding their passage, stick in the public consciousness as a statute of great importance;²³ and third, the understanding of constitutional statutes as statutes that operate to regulate the institutions and powers of the State.²⁴ The purpose of this assessment is not to add to the definitional debates about the nature of ‘constitutional’ statutes, but to provide a basis to assess whether the *RDA* could be considered a ‘constitutional’ statute.

A Fundamental Values or Rights

A statute can be designated as ‘constitutional’ because of its subject matter.²⁵ Commonly this is because the subject matter relates to fundamental values or rights.²⁶ This focus on fundamental values reflects an understanding of a constitution as sitting at the top of a normative pyramid that establishes and articulates the social values of a society.²⁷ Consequently, ‘constitutional’ statutes are statutes that are designed to guide human behaviour and shape the character of the State.²⁸ As such, ‘constitutional’ statutes articulate the State’s fundamental social aspirations and values.²⁹ In the UK context, in *Thoburn v Sunderland City Council* Laws LJ outlined

¹⁶ Vanessa MacDonnell, ‘A Theory of Quasi-Constitutional Legislation’ (2016) 53(2) *Osgoode Hall Law Journal* 508, 509–10.

¹⁷ Helis (n 7) 4.

¹⁸ *Canadian Bill of Rights*, SC 1960, c 44.

¹⁹ *Ibid.*

²⁰ Stephenson (n 6) 27–8.

²¹ Ahmed and Perry (n 15) 465; MacDonnell (n 16) 510.

²² *Thoburn v Sunderland City Council* [2003] QB 151, 186–87 [62] (Laws LJ) (‘*Thoburn*’).

²³ William N Eskridge Jr and John Ferejohn, ‘Super-Statutes’ (2001) 50(5) *Duke Law Journal* 1215, 1216. This idea is drawn and adapted from Bruce Ackerman, *We the People (Vol 2): Transformations* (Harvard University Press, 1998) 4–5.

²⁴ David Feldman, ‘The Nature and Significance of Constitutional Legislation’ (2013) 129 (July) *Law Quarterly Review* 343, 357.

²⁵ Helis (n 7) 4–10.

²⁶ *Ibid.*

²⁷ Aharon Barak, *Purposive Interpretation in Law* (Princeton University Press, 2011) 190.

²⁸ *Ibid.*

²⁹ *Ibid.*

two forms of legislation that qualified as ‘constitutional’ legislation.³⁰ The first form of ‘constitutional’ legislation is legislation that outlines the conditions of the legal relationship between the citizen and the State in a general, overarching manner. The second is legislation that either enlarges or diminishes the scope of what are ‘fundamental constitutional rights.’³¹ The Supreme Court of the United Kingdom has also accepted the existence of constitutional statutes.³² This second category, in particular, seems to be drawing on the idea that statutes can be ‘constitutional’ where they draw on fundamental rights and values.

A number of constitutional commentators from the UK including Feldman dispute the categorical distinction described by Laws LJ.³³ Feldman criticises Laws LJ’s categorisation because the categories do not have clear and distinct boundaries.³⁴ Feldman argues that many kinds of legislation change the legal relationship between the citizen and State, and further, will change over time because fundamental constitutional values are not fixed.³⁵

The idea that Commonwealth discrimination legislation invokes fundamental values has been raised before. Saunders and Donaldson describe the introduction of discrimination law as implementing the value of equality in the Australian legal system.³⁶ Evans lists anti-discrimination legislation as one form of ‘constitutional’ legislation in Australia, drawing on the definition provided in *Thoburn*.³⁷ In this way, the *RDA*, as well as the other Commonwealth discrimination laws, can have a kind of ‘constitutional’ status based on the understanding that discrimination law is designed to give legal force to the fundamental values of non-discrimination and equality.³⁸ The *RDA*, in particular, does so by enlarging a person’s rights by giving them the capacity to challenge laws and actions that are based on irrelevant considerations, such as race.³⁹ As a consequence, the *RDA* and other discrimination legislation have the normative force necessary for a ‘constitutional’ statute.

B *Impact on Political Norms and Structures*

A constitutional statute can also be understood as akin to what Eskridge Jr and Ferejohn describe as a ‘super-statute’.⁴⁰ A ‘super-statute’ is a statute that establishes a new normative or institutional framework for state practice. Eskridge Jr and Ferejohn describe a super-statute as a statute that embodies a significant normative principle that is adopted after an intense political struggle. Such a statute

³⁰ *Thoburn* (n 22).

³¹ *Ibid* 186–7 [62]–[64].

³² *H v Lord Advocate* [2013] 1 AC 413, 435–6 [30] (Lord Hope DPSC); *R (HS2 Action Alliance Ltd) v Secretary of State for Transport* [2014] 1 WLR 324, 382–3 [208] (Lord Neuberger PSC and Lord Mance JSC).

³³ Feldman, ‘The Nature and Significance of Constitutional Legislation’ (n 24) 353–4.

³⁴ *Ibid* 346–8.

³⁵ *Ibid* 346. See also Ahmed and Perry (n 15) 464.

³⁶ Saunders and Donaldson (n 5) 36.

³⁷ Evans (n 13) 120.

³⁸ Gaze, ‘Racial Discrimination Act’ (n 10) 178.

³⁹ Through its prohibitions on discrimination in *RDA* (n 8) ss 9–17.

⁴⁰ Eskridge Jr and Ferejohn (n 23) 1216.

‘pervasively affects executive and legislative action’.⁴¹ Such a statute ‘sticks’ in public culture in a way that the values embedded in the statute have a broad effect on the development of law.⁴² Eskridge Jr and Ferejohn point to the *Civil Rights Act of 1964* as an illustration of the kind of statute that they consider is representative of a ‘super-statute’.⁴³ A statute’s status as a super-statute is not embedded at the time of passage, but instead is developed over time through a series of contestations surrounding its importance and meaning.⁴⁴ In this conceptualisation, constitutionalism is the culmination of a contest about fundamental values and rights. A constitution or ‘super-statute’ is the outcome of such a struggle.⁴⁵

The *RDA*’s path to passage and its place in the public consciousness since passage also demonstrates the existence of a ‘battle over fundamental rights and values’ described by Ackerman,⁴⁶ and Eskridge Jr and Ferejohn.⁴⁷ The *RDA* was the first statute in Australia prohibiting discrimination at the Commonwealth level. In introducing the 1975 Bill, Attorney-General Kep Enderby acknowledged that the purpose of the legislation was to implement Australia’s obligations contained in the *International Convention on the Elimination of All Forms of Racial Discrimination*,⁴⁸ to remedy the inadequacies of the common law and to educate the public about the ‘undesirable and unsocial consequences of discrimination ... and make them more obvious and conspicuous’.⁴⁹

There are competing perspectives on whether the *RDA* was the culmination of a struggle for civil rights. The implication of these competing perspectives is that the passage of the *RDA* does not fit neatly within Ackerman’s understanding of constitutional statutes as the culmination of a political struggle.⁵⁰ On the one hand, the legislation had bipartisan support, and much of the academic commentary from the time, though critical of the *RDA*’s limitations, does not indicate a significant level of controversy about the notion of protection from discrimination.⁵¹ Further, in his analysis of Australians’ opinions on race from 1943 onwards, Markus found that there was no specific polling data on the introduction of the *RDA* and that its introduction and passage did not result in public controversy.⁵²

⁴¹ Ibid.

⁴² Ibid.

⁴³ Ibid 1237, citing *Civil Rights Act of 1964*, Pub L No 88-352, 78 Stat 241.

⁴⁴ Ibid 1216–7.

⁴⁵ Ackerman (n 23) 5, 170.

⁴⁶ Ibid 170.

⁴⁷ Eskridge Jr and Ferejohn (n 23) 1216.

⁴⁸ *International Convention on the Elimination of All Forms of Racial Discrimination*, opened for signature 21 December 1965, 660 UNTS 195 (entered into force 4 January 1969) (‘ICERD’).

⁴⁹ Commonwealth, *Parliamentary Debates*, House of Representatives, 13 February 1975, 285 (Kep Enderby, Attorney-General).

⁵⁰ Ackerman (n 23) 170.

⁵¹ See, eg, David Partlett, ‘The *Racial Discrimination Act 1975* and the *Anti-Discrimination Act 1977*: Aspects and Proposals for Change’ (1977) 2(2) *UNSW Law Journal* 152; David Partlett, ‘Benign Racial Discrimination: Equality and Aborigines’ (1979) 10(3) *Federal Law Review* 238; Brian Kelsey, ‘A Radical Approach to the Elimination of Racial Discrimination’ (1975) 1(1) *UNSW Law Journal* 56.

⁵² Andrew Markus, ‘Australian Opinion on Issues of Race: A Broad Reading of Opinion Polls, 1943–2014’, *Perspectives on the Racial Discrimination Act: Papers from the 40 Years of the Racial Discrimination Act Conference* (Australian Human Rights Commission, 2015) 19–20.

Lino places the *RDA* within the context of a broader effort for Indigenous constitutional recognition.⁵³ He argues that the *RDA* owes its very existence, in significant part, to Indigenous activism focused on the efforts to end the discriminatory legal regime that operated in Queensland throughout the 1970s.⁵⁴ In his view, the enactment of the *RDA* served both to provide a capacity to curtail the operation of the racist legal regime in Queensland and to recognise Indigenous peoplehood in a symbolic sense.⁵⁵ As a piece of ‘ordinary’ legislation, Lino acknowledges that the *RDA* lacks ‘big-C’ constitutional force.⁵⁶ He nevertheless argues it forms part of the ‘small-C’ constitutional recognition of Indigenous peoples through securing protection from racial discrimination for Indigenous persons at a national level.⁵⁷ From this perspective, the *RDA* has had, since its inception, the kind of symbolic status required for a form of constitutional legislation.

The debates surrounding the *RDA* since its inception also show its ‘stickiness’ in the battle over Australian public values.⁵⁸ This is particularly demonstrated through the prolonged efforts to amend s 18C of the *RDA* in light of the 2011 decision in *Eatock v Bolt*.⁵⁹ Despite sustained efforts to remove or significantly limit the application of the racial vilification provisions in the *RDA* over a number of years,⁶⁰ these have had limited success, demonstrating the force and effectiveness of the *RDA* as a significant and important piece of human rights legislation in public consciousness. Consequently, this ‘stickiness’ in the public consciousness and the difficulties faced by legislatures in narrowing the *RDA*’s terms with respect to s 18C appears to show commonalities with what Eskridge Jr and Ferejohn describe as a super-statute.

C *Regulating a Fundamental Feature of the Lawmaking Process*

Instead of a focus on either the subject matter or the public, political and legal debates surrounding a statute, other definitions focus on the form and function of the statute.⁶¹ Preferring a definition focused on the function of the statute, Feldman considers that statutes are ‘constitutional’ where they establish and regulate the various institutions of the State.⁶² He considers that the institutional approach to ‘constitutional’ statutes is most in keeping with the core function of a constitution. ‘Constitutional’ statutes are therefore statutes that establish institutions and confer

⁵³ Dylan Lino, *Constitutional Recognition: First Peoples and the Australia Settler State* (Federation Press, 2018) ch 6.

⁵⁴ *Ibid* 174.

⁵⁵ *Ibid* 177.

⁵⁶ *Ibid* 187–8.

⁵⁷ *Ibid* 188.

⁵⁸ Eskridge Jr and Ferejohn (n 23) 1230–1.

⁵⁹ *Eatock v Bolt* (2011) 197 FCR 261. See also Williams and Reynolds (n 3) 242.

⁶⁰ For a discussion of this debate, see Adrienne Stone, ‘The Ironic Aftermath of *Eatock v Bolt*’ (2015) 38(3) *Melbourne University Law Review* 926, 936–45; Katharine Gelber and Luke McNamara, ‘Freedom of Speech and Racial Vilification in Australia: “The Bolt Case” in Public Discourse’ (2013) 48(4) *Australian Journal of Political Science* 470; Amanda Porter, ‘Words Can Never Hurt Me? Sticks, Stones and Section 18C’ (2015) 40(2) *Alternative Law Journal* 86.

⁶¹ Feldman, ‘The Nature and Significance of Constitutional Legislation’ (n 24) 351.

⁶² David Feldman, ‘Statutory Interpretation and Constitutional Legislation’ (2014) 130 (July) *Law Quarterly Review* 473, 473.

and confine functions, responsibilities and powers on such institutions.⁶³ Feldman concludes that this is a preferable definition because such a definition demarcates statutes that state *what the law is* from those that define *how a law is made*.⁶⁴ While drawing a distinction between constitutional statutes in jurisdictions with capital-C constitutions and jurisdictions without a capital-C constitution, Stephenson also defines constitutional statutes as those that regulate a fundamental feature of the lawmaking process in terms of the enactment, administration and interpretation of the law.⁶⁵ The regulation of the lawmaking process can involve both the making of laws by the legislature, as well as the administration of the laws by the executive and the interpretation of laws by the judiciary.⁶⁶ What differentiates a ‘constitutional’ statute from a capital-C constitution is that these statutes lack some of the attributes of capital-C-style constitutions with respect to entrenchment and superiority.⁶⁷

The *RDA* could be understood as ‘constitutional’ legislation because of its capacity to constrain all state legislative and executive power.⁶⁸ Its capacity to do so goes beyond that which all Commonwealth legislation has the power to do by virtue of s 109 of the *Australian Constitution*. Of particular importance is s 10 of the *RDA*, which provides a right to equality before the law.⁶⁹ As highlighted in the introduction, s 10 is distinctive in the discrimination law landscape, with no equivalent provisions in other Commonwealth or state discrimination legislation, which generally have a horizontal effect. Section 10 operates in two ways. It can invalidate state laws that would otherwise operate in a discriminatory manner on the basis of race.⁷⁰ Alternatively, where a state law omits to make a right universal, s 10 confers that right on to persons of a particular race.⁷¹ It is the capacity of s 10 of the *RDA* to invalidate state laws that makes it unique in the Australian discrimination landscape.⁷² While not focused on discriminatory laws, but discriminatory conduct, s 9 of the *RDA* has also been used to challenge conduct by state executive officers such as police officers in the carrying out of their duties in a racially discriminatory fashion.⁷³

Williams and Reynolds identify that the ‘constitutional’ value of discrimination legislation is demonstrated primarily through the supremacy of federal legislation over state legislation.⁷⁴ This supremacy enables federal discrimination legislation to set standards of conduct at both a federal and state

⁶³ Feldman, ‘The Nature and Significance of Constitutional Legislation’ (n 24) 350.

⁶⁴ Ibid.

⁶⁵ Stephenson (n 6) 28.

⁶⁶ Ibid 29.

⁶⁷ Ibid 28.

⁶⁸ Williams and Reynolds (n 3) 242–3.

⁶⁹ *RDA* (n 8) s 10.

⁷⁰ *Gerhardy v Brown* (1985) 159 CLR 70, 98 (Mason CJ) (‘*Gerhardy*’).

⁷¹ Ibid. See also *Western Australia v Ward* (2002) 213 CLR 1, 99–100 (Gleeson CJ, Gaudron, Gummow and Hayne JJ) (‘*Ward*’).

⁷² Neil Rees, Simon Rice and Dominique Allen, *Australian Anti-Discrimination & Equal Opportunity Law* (Federation Press, 2018) 264.

⁷³ See, eg, *Wotton v Queensland (No 5)*, a class action claim against the Queensland Government on behalf Aboriginal and Torres Strait Islanders living on Palm Island, which found that the police had unlawfully discriminated against residents: (2016) 352 ALR 146 (‘*Wotton*’).

⁷⁴ Williams and Reynolds (n 3) 246.

level.⁷⁵ Stephenson also adds other discrimination statutes related to sex, age and disability to the list of Australian ‘constitutional’ statutes.⁷⁶ This is on the basis that discrimination and other human rights legislation function to regulate a fundamental feature of the lawmaking process. This includes not only the enactment of laws by the legislature, but also the actions of the executive and the interpretation of laws by the judiciary.⁷⁷ This definition would appear not only to include s 10 of the *RDA*, but also ss 9 and 9(1A) through the capacity to curtail executive action.⁷⁸ An example of the use of s 9 in this respect is the case of *Wotton v Queensland (No 5)*, in which it was successfully argued that the conduct of Queensland police after the death of an Aboriginal Australian man in custody was racially discriminatory against members of the Aboriginal community.⁷⁹

Though the *RDA* has not been declared ‘constitutional’, in some cases of inconsistency, courts have applied similar principles to those applied in jurisdictions that recognise constitutional or quasi-constitutional statutes. In other jurisdictions, this recognitional status can have significant effects on the relationship of that statute with other ‘ordinary’ statutes. In particular, where there is an inconsistency of terms, the ordinary statute is read-down to ensure consistency with the fundamental rights contained in the ‘constitutional’ statute, providing a derogation from the general principle of *generalia specialibus non derogant*, or that specific legislation has precedence over general legislation.⁸⁰

Further, the principle of implied repeal does not extend to constitutional statutes, and the legislature must use clear and unambiguous language where a later statute is interpreted to repeal sections of an earlier constitutional statute.⁸¹ In the UK, Lord Neuberger and Lord Justice Sales have both remarked extra-curially that this limitation of the rule of implied repeal is akin to applying the principle of legality to fundamental common law rights.⁸² In considering the classification of ‘constitutional’ statutes in the UK, and its possible application in Australia, French CJ also appeared to share this understanding as to the effect of the limitation on the principle of implied repeal.⁸³

In the Canadian context, the interpretive rules that are applied to quasi-constitutional legislation have had a significant impact in expanding the scope and application of human rights principles.⁸⁴ The primacy of ‘constitutional’ legislation is recognised by Canadian courts based on the ‘fundamental’ character of

⁷⁵ Ibid.

⁷⁶ Stephenson (n 6) 35.

⁷⁷ Ibid 29.

⁷⁸ *RDA* (n 8) ss 9, 9(1A), 10.

⁷⁹ *Wotton* (n 73).

⁸⁰ Ahmed and Perry (n 15) 463.

⁸¹ Ibid 462.

⁸² Lord Neuberger, ‘The Constitutional Role of the Supreme Court in the Context of Devolution in the UK’ (Lord Rodger Memorial Lecture 2016, Glasgow, 14 October 2016) [18]–[20]; Lord Justice Philip Sales, ‘Rights and Fundamental Rights in English Law’ (2016) 75(1) *Cambridge Law Journal* 86, 91.

⁸³ *Cadia* (n 11) 217–18.

⁸⁴ Claire Mummé, ‘At the Crossroads in Discrimination Law: How the Human Rights Codes Overtook the *Charter* in Canadian Government Services Cases’ (2012) 9(1) *Journal of Law & Equality* 103, 106.

constitutional legislation.⁸⁵ Primacy requires that courts adopt a two-stage process when considering inconsistency between a constitutional statute and an ordinary statute.⁸⁶ First, the courts attempt to read the ordinary statute in a manner that is consistent with the fundamental rights contained within the quasi-constitutional statute.⁸⁷ Helis describes this aspect of primacy as similar to the ‘weak-form’ model of constitutionalism and judicial review in some other Commonwealth jurisdictions.⁸⁸ At the second stage, where a conflict cannot be resolved, the ordinary statute is inoperable to the extent of the inconsistency.⁸⁹ The two-stage test was developed pursuant to the *Canadian Bill of Rights*.⁹⁰ The Supreme Court of Canada has applied this test to the statutory human rights codes (which are equivalent to Australian anti-discrimination legislation) since its decision in *Insurance Corporation of British Columbia v Heerspink*.⁹¹ This rule applies regardless of whether the statutory human rights codes have an explicit primacy clause or not.⁹²

As Chen has recognised, in the context of statutory interpretation more generally, the commentary and Australian jurisprudence recognise that the rule of implied repeal is significantly limited.⁹³ It is limited because courts start from the proposition that statutes do not contradict each other. Thus, courts seek to apply a principle of harmonious construction so that both statutes can continue to operate unimpeded.⁹⁴ Nevertheless, these principles of interpretation with respect to inconsistency and the limitation of implied repeal are evident in the Australian case law on the *RDA*.

In interactions or inconsistencies between the *RDA* and state legislation, due to s 109 of the *Australian Constitution*, the supremacy of the *RDA* over state legislation is clear.⁹⁵ While other Commonwealth discrimination Acts can render state schemes that are discriminatory on the basis of sex, disability or age inoperable to the extent of their inconsistency with Commonwealth legislation, the incorporation of s 10 of the *RDA* provides a particularly compelling force in this respect. Section 10 applies to both state and federal laws. The effect of s 10 with respect to state legislation, as was first articulated by Mason J in *Gerhardy v Brown*, is two-fold.⁹⁶ Section 10 can invalidate state laws that would otherwise operate in a discriminatory manner on the basis of race.⁹⁷ Alternatively, where a state law fails to make a right universal, s 10 confers that right on to persons of a particular race.⁹⁸

⁸⁵ *Insurance Corporation of British Columbia v Heerspink* [1982] 2 SCR 145, 157–8 (Lamer CJ) (*‘Heerspink’*).

⁸⁶ Helis (n 7) 91–2.

⁸⁷ Ibid 92–3.

⁸⁸ Ibid.

⁸⁹ Ibid 94.

⁹⁰ *Canadian Bill of Rights* (n 18). See also *R v Drybones* [1970] SCR 282, 305.

⁹¹ *Heerspink* (n 85) 157–8 (Lamer CJ).

⁹² Helis (n 7) 91–2.

⁹³ Bruce Chen, ‘The Principle of Legality: Protecting Statutory Rights from Statutory Infringement?’ (2019) 41(1) *Sydney Law Review* 73, 89.

⁹⁴ Ibid.

⁹⁵ Williams and Reynolds (n 3) 243. See also *Viskauskas v Niland* (1983) 153 CLR 280 (*‘Viskauskas’*); *University of Wollongong v Metwally* (1983) 158 CLR 447 (*‘Metwally’*).

⁹⁶ *Gerhardy* (n 70) 98–9.

⁹⁷ Ibid 98.

⁹⁸ Ibid. See also *Ward* (n 71) 99–100 (Gleeson CJ, Gaudron, Gummow and Hayne JJ).

Mason J's interpretation of the effect and application of s 10 was later accepted by a majority of the High Court in *Western Australia v Ward*⁹⁹ and in the judgments of French CJ, Hayne J and Bell J in *Maloney v The Queen*.¹⁰⁰

Further, while in the Commonwealth context the *RDA* cannot override other Commonwealth legislation, the Full Court of the Federal Court of Australia recognised in *Vanstone v Clarke* the need to interpret other Commonwealth legislation consistently with the terms of the *RDA*.¹⁰¹ In particular, Commonwealth legislation is to be interpreted consistently with the terms of s 10 to ensure harmonious operation.¹⁰² Despite this, the case law demonstrates limited success for claimants seeking to challenge both state and Commonwealth legislation on the basis of inconsistency with s 10 of the *RDA*. With respect to Commonwealth legislation that is potentially inconsistent with the *RDA*, while such an argument has been raised in 11 cases,¹⁰³ there is only one decision in which the inconsistency argument was successful: *Shi v Minister for Immigration and Citizenship*.¹⁰⁴

The focus in *Shi* was s 499(1) of the *Migration Act 1958* (Cth).¹⁰⁵ Section 499(1) empowered the Minister to give directions as to how to exercise the discretion to cancel a visa pursuant to s 501. In one of these directions, 'Direction No 401', a delegate was instructed to give favourable consideration to 'ties and linkages' to Australia.¹⁰⁶ In reviewing and approving the decision to cancel the appellant's visa, the tribunal member focused considerable attention on the fact that 'a large part of his upbringing and character formation was in China'.¹⁰⁷ Perram J concluded that such a determination could not be made pursuant to s 499(1) of the *Migration Act 1958* (Cth) because it was inconsistent with the terms of s 10 of the *RDA*. Consequently, the power granted to the Minister pursuant to s 499(1) had to be interpreted consistently with s 10 of the *RDA*, limiting the scope of operation s 499(1).¹⁰⁸ Interpretation of s 499(1) consistent with the *RDA* was justified on the basis that 'it would require express words to convey an intention that a general power ... authorised the repository to repeal or amend Parliament's own enactments'.¹⁰⁹

One challenge for potential claimants has been that where there is potential inconsistency between different federal statutes, the legislature can choose to

⁹⁹ *Ward* (n 71) 99 (Gleeson CJ, Gaudron, Gummow and Hayne JJ).

¹⁰⁰ *Maloney v The Queen* (2013) 252 CLR 168, 179 (French CJ) 200–1 (Hayne J), 244 (Bell J) ('*Maloney*').

¹⁰¹ *Vanstone v Clarke* (2005) 147 FCR 299, 352–4 (Weinberg J).

¹⁰² *Ibid.*

¹⁰³ *Vanstone v Clark* (n 101); *Macabenta v Minister for Immigration and Multicultural Affairs* (1998) 90 FCR 202; *Nguyen v Refugee Review Tribunal* (1997) 74 FCR 311; *Department of Veterans' Affairs v P* (1998) 79 FCR 594; *Yildiz v Minister for Immigration and Ethnic Affairs* (1982) 70 FLR 105; *Sremcevic v Gurry* (1994) 51 FCR 194; *Trau v Repatriation Commission* (1998) 88 FCR 349; *Shi v Minister for Immigration and Citizenship* (2011) 123 ALD 46 ('*Shi*'); *Melkman v Federal Commissioner of Taxation* (1988) 20 FCR 331; *Sahak v Minister for Immigration and Multicultural Affairs* (2002) 123 FCR 514; *Naen v Minister for Immigration and Multicultural and Indigenous Affairs* [2003] FCA 216.

¹⁰⁴ *Shi* (n 103).

¹⁰⁵ *Ibid.*

¹⁰⁶ *Ibid* 47 [5].

¹⁰⁷ *Ibid* 47 [6].

¹⁰⁸ *Ibid* 50 [20]–[21].

¹⁰⁹ *Ibid* 50 [20], quoting *De L v Director-General, NSW Department of Community Services (No 2)* (1997) 190 CLR 207, 212 (Brennan CJ and Dawson J).

exclude the operation of the *RDA* with express and clear legislative language, rather than potentially risk a challenge to the validity of the legislation.¹¹⁰ A well-known example of this type of legislative action was taken with respect to the intervention in the Northern Territory ('NT'). The Commonwealth declared that the package of measures adopted with respect to the intervention were 'special measures' pursuant to s 8 of the *RDA* and explicitly excluded the operation of pt II of the *RDA*.¹¹¹ Pt II of the *RDA* prohibits unlawful discrimination. The possibility of such an express exclusion of operability clearly limits the capacity of the *RDA* to facilitate widespread social change.¹¹² However, the fact that the legislature considered there was a need to exclude pt II of the *RDA* points to a degree of 'specialness' that the *RDA* nevertheless holds. Without such 'specialness', it is difficult to understand why the traditional rules of statutory interpretation would not apply; namely, that the older and more general statute, the *RDA*, would be read-down to ensure consistency with the newer and more specific legislation pertaining to the intervention.

In the context of state legislation, there have been some important and highly significant successful challenges to state legislation. The *RDA* has been crucial in limiting the curtailment of native title rights in cases such as *Mabo v Queensland (No 1)*,¹¹³ the *Second Native Title Act Case*,¹¹⁴ *Ward*,¹¹⁵ *Jango v Northern Territory*,¹¹⁶ and *James v Western Australia*.¹¹⁷ However, the *RDA*'s effectiveness in other spaces has been significantly limited. Outside the area of native title, courts have only accepted claims of inconsistency where state discrimination laws were more progressive or expansive than the *RDA* in protecting non-discrimination rights.¹¹⁸

Though numerous claims have attempted to nullify potentially discriminatory state-based laws, few have been successful.¹¹⁹ As Williams and Reynolds have identified, there are three reasons such claims have not succeeded.¹²⁰ First, claims failed where the court concluded that the state legislation does not discriminate on the basis of race.¹²¹ Second, claims failed where they could not demonstrate that a

¹¹⁰ Jonathon Hunyor, 'Is it Time to Re-Think Special Measures under the *Racial Discrimination Act*? The Case of the Northern Territory Intervention' (2009) 14(2) *Australian Journal of Human Rights* 39, 61.

¹¹¹ *Northern Territory National Emergency Response Act 2007* (Cth) ss 132(1)–(2); *Social Security and Other Legislation Amendment (Welfare Payment Reform) Act 2007* (Cth) ss 4(2)–(5); *Families, Community Services and Indigenous Affairs and Other Legislation Amendment (Northern Territory National Emergency Response and Other Measures) Act 2007* (Cth) ss 4(1)–(2).

¹¹² A point that has been made before, see Hunyor (n 110); Lino 'Thinking outside the *Constitution*' (n 4).

¹¹³ *Mabo v Queensland (No 1)* (1988) 166 CLR 186 ('*Mabo (No 1)*').

¹¹⁴ *Western Australia v Commonwealth* (1995) 183 CLR 373 ('*Second Native Title Act Case*').

¹¹⁵ *Ward* (n 71).

¹¹⁶ *Jango v Northern Territory* (2006) 152 FCR 150.

¹¹⁷ *James v Western Australia* (2010) 184 FCR 582.

¹¹⁸ See, eg, *Viskauskas* (n 95); *Metwally* (n 95); *Central Northern Adelaide Health Service v Atkinson* (2008) 103 SASR 89.

¹¹⁹ Williams and Reynolds (n 3) 243.

¹²⁰ *Ibid* 248.

¹²¹ *Ibid*. The example given for this reason for failure was *Aurukun Shire Council v Chief Executive Officer, Office of Liquor, Gaming and Racing in the Department of Treasury* [2012] 1 Qd R 1 ('*Aurukun*'). I would suggest that also with this category is *Bropho v Western Australia* (2008) 169 FCR 59.

‘fundamental right or freedom’ had been infringed.¹²² Third, claims failed where courts accepted that though a provision may operate in a discriminatory manner, it is nevertheless justified and has been ‘saved’ as a special measure.¹²³ Though these special interpretive rules with respect to the limitation of the principle of implied repeal and the primacy of the ‘constitutional’ statute can apply to the *RDA*, there is nevertheless still a lack of success for claimants when making these arguments.

D *Can the Perceived ‘Specialness’ of the Racial Discrimination Act 1975 (Cth) be Considered ‘Constitutional’?*

Part of the difficulty in determining if the *RDA* is ‘constitutional’ in a definitional sense is the malleability of many of the definitions provided. As Feldman identified, the problem with the definition provided in *Thoburn* is that the list of ‘fundamental’ rights is not a closed list, allowing for a degree of flexibility as to what can be added to such a list.¹²⁴ Similarly, political and legal struggles can emerge in a wide variety of areas of regulation — thus, again, on its own, such struggles would appear to be insufficient to denote any kind of special status. Further, in Australia, the nature of federalism necessarily means that Commonwealth statutes, provided they have a sufficient constitutional head of power, may intrude into areas of state responsibility. If one only considered the impact on legislation-making powers, one could conclude that a wide array of statutes are ‘constitutional’. Consequently, applying any of these definitions in the Australian context would appear to be insufficient on their own. Instead, it would seem that an amalgamation of these factors — the normative content, the symbolic and political nature of the legislation, and its role in regulating the lawmaking process — would be more consistent with the Australian experience.

The degree of flexibility attached to each possible definition of ‘constitutional’ legislation, as well as their specific application to the Australian context, ultimately means that any ‘constitutional’ force of the *RDA* comes from a combination of factors. These include the constraining influence of Commonwealth law on state legislation and the administration of public services by state governments, as well as the statute’s subject matter and its importance in the public consciousness. On their own, these identifying features do not provide significant ‘constitutional’ force. Instead, it is the combination of these factors that justify the inclusion of discrimination legislation as ‘constitutional’.

III The Interpretation of ‘Constitutional’ Statutes and the *Racial Discrimination Act 1975 (Cth)*

In other jurisdictions that recognise constitutional statutes, there is some divergence as to whether the constitutional status of a statute influences the interpretation of its terms. In this section, I briefly outline the different approaches adopted in Canada

¹²² Williams and Reynolds (n 3) 248–9. The example given for the reason for failure was also *Aurukun* (n 121). I suggest that this category also includes *Morton v Queensland Police Service* (2010) 240 FLR 269.

¹²³ Williams and Reynolds (n 3) 248. The authors utilise the example of *Maloney* (n 100) as an example of this kind of case. Similarly, *Gerhardy* (n 70) fits within this category.

¹²⁴ Feldman, ‘The Nature and Significance of Constitutional Legislation’ (n 24) 346.

and the UK before considering the interpretive approach adopted in Australia in the context of the *RDA*. I focus particularly on the idea that constitutional statutes should be interpreted in a fashion that is ‘broad, liberal and purposive’.¹²⁵ The above section highlighted that utilising a variety of definitions, the *RDA* can be considered ‘constitutional’. The legislature and the broader community appear to perceive the *RDA* to have a special status and, at times, the *RDA*’s terms have been applied where there is inconsistency between statutes, not only due to inconsistency within s 109 of the *Australian Constitution*, but also where there is inconsistency between Commonwealth statutes. However, my analysis of s 10 jurisprudence in this section will demonstrate that though the interpretation of the *RDA* could be considered ‘broad’, in contrast to other jurisdictions it is not necessarily liberal or purposive.

A Principles of Interpretation and ‘Constitutional’ Statutes

In Canada, Helis has observed that a defining characteristic of Canadian quasi-constitutional law is the manner in which the substantive provisions are interpreted.¹²⁶ The Supreme Court of Canada interprets ‘constitutional’ legislation differently from other forms of legislation because ‘it is inappropriate to rely solely on a strictly grammatical analysis, particularly with respect to the interpretation of legislation which is constitutional or quasi-constitutional in nature’.¹²⁷

This has resulted in an interpretive style which is ‘broad, liberal and purposive’ in nature and in practice significantly expands the rights that are contained in quasi-constitutional legislation.¹²⁸ Where the Canadian Supreme Court has accepted that legislation is quasi-constitutional, this has led to protective provisions being interpreted broadly, in keeping with the statute’s rights-protective quality.¹²⁹ Conversely, the defences and justifications are read down to narrow their possible effect.¹³⁰

The importance of a broad, liberal and purposive interpretation on the reach and effect of quasi-constitutional statutes is demonstrated through a contrast between the effectiveness of the *Canadian Bill of Rights* and the human rights codes.¹³¹ Due to the restrictive nature of the interpretation of the *Canadian Bill of Rights*, it was relatively weak in securing any significant change because it was rare that other statutes were inconsistent with its provisions.¹³² In contrast, the broad, liberal and purposive approach to interpretation of the human rights codes, combined with primacy of the human rights codes over other statutes, have allowed for an expansive protection of human rights.¹³³

¹²⁵ Helis (n 7) 3.

¹²⁶ Ibid 17.

¹²⁷ *Quebec (Commission des Droits de la Personne et des Droits de la Jeunesse) v Montréal (City)* [2000] 1 SCR 665, 684 [30] (citations omitted).

¹²⁸ Helis (n 7) 17.

¹²⁹ Ibid 23.

¹³⁰ Ibid 78.

¹³¹ Ibid 230–1.

¹³² Ibid 215.

¹³³ Ibid 237.

While Canadian human rights statutes are statutes with horizontal effect and a focus on discrimination in employment and the provision of goods and services, their reach is expansive, particularly in the context of government services.¹³⁴ Human rights statutes were designated by the Canadian Supreme Court as quasi-constitutional legislation in the 1980s.¹³⁵ The effect of this on the interpretation of the non-discrimination rights contained in the legislation has been significant. As Vizkelely has articulated regarding the expanded scope of statutory discrimination law:

Interestingly, the breakthrough has come as a result of judicial, not legislative, intervention ... As for the courts, there has been an ever-growing tendency especially at the higher levels, to recognize the special nature of human rights legislation. Fortified, perhaps by their heightened responsibilities under the *Canadian Charter of Rights and Freedoms*, courts have shown that they are now prepared to look beyond the narrow and literal constructions of anti-discrimination laws and to give effect to their purpose. It is on the basis of a liberal approach such as this that courts have recognized the effects of the concept of discrimination.¹³⁶

In practice, this has meant that it was through judicial, rather the legislative, intervention that Canada's human rights regime began to incorporate progressive concepts. The Supreme Court expanded the definition of discrimination to include indirect discrimination (where a provision or practice may be non-discriminatory on its face, but is discriminatory in its effect)¹³⁷ and to require reasonable adjustments for all grounds of discrimination.¹³⁸ In contrast, the Canadian courts have adopted a narrow interpretation of 'bona fide' occupational requirements (in the case of employment) and 'bona fide and reasonable' requirements (in the case of other areas of operation).¹³⁹

In contrast to the Canadian position, in the UK, there is no 'special' interpretive style adopted with respect to 'constitutional' statutes. As the UK Supreme Court explained in *Imperial Tobacco Ltd v Lord Advocate (Scotland)*, the *Scotland Act 1998* (Scot) was to be interpreted in the same way as one would interpret any other UK statute.¹⁴⁰ This was reaffirmed by Lord Hope in *Attorney General v National Assembly for Wales Commission* with whom the other justices agreed.¹⁴¹ However, in his judgment, Lord Hope acknowledged that:

¹³⁴ Mummé (n 84) 106.

¹³⁵ *Heerspink* (n 85) 157–8 (Lamer CJ).

¹³⁶ Béatrice Vizkelely, *Proving Discrimination in Canada* (Carswell, 1987) 237–8.

¹³⁷ *Ontario Human Rights Commission v Simpsons-Sears* [1985] 2 SCR 536.

¹³⁸ *Bhinder v CN* [1985] 2 SCR 561.

¹³⁹ *British Columbia (Public Service Employee Relations Commission) v BCGSEU* [1999] 3 SCR 3; *British Columbia (Superintendent of Motor Vehicles) v British Columbia (Council of Human Rights)* [1999] 3 SCR 868. It should be noted here that the concept of discrimination is unified in Canada (with no distinction between direct and indirect discrimination). Thus, in all claims Canadian courts consider whether the discrimination is justified — as compared to Australian claims, for which the reasonableness is only assessed in indirect claims.

¹⁴⁰ *Imperial Tobacco Ltd v Lord Advocate (Scotland)* 2013 SCLR 121.

¹⁴¹ *Attorney General v National Assembly for Wales Commission* [2013] 1 AC 792. Prior to this decision, it was thought that 'constitutional' legislation was to be interpreted 'generously' and 'purposively' due to the decision in *Robinson v Secretary of State for Northern Ireland* [2002] NI 390. See also Stephen J Dimelow, 'The Interpretation of "Constitutional" Statutes' (2013) 129 (October) *Law Quarterly Review* 498, 498; David Feldman, 'Statutory Interpretation and Constitutional Legislation' (n 62) 497.

The rules to which the court must apply in order to give effect to it are those laid down by the statute, and the statute must be interpreted like any other statute. But the purpose of the Act has informed the statutory language, and it is proper to have regard to it if help is needed as to what the words mean.¹⁴²

Drawing on this statement, Feldman argues that while the interpretation of ‘constitutional’ statutes is no different from ‘ordinary’ statutes, by their very nature, the purposes of ‘constitutional’ statutes are highly general.¹⁴³ They are general in nature because they need to apply in a variety of circumstances and introduce general rules. Consequently, such statutes must be understood with a high level of generality and refer to the broader overarching values that such a statute embodies.¹⁴⁴ Thus, though in the UK context, the process of interpretation is not special or different, it nevertheless requires judges to draw on more general and possibly ‘aspirational’ values in interpreting constitutional legislation.¹⁴⁵

B *The Interpretation of s 10 of the Racial Discrimination Act 1975 (Cth)*

While Australian courts have been willing to adopt special interpretive rules for limitation of the principle of implied repeal and to give the *RDA* primacy in some cases with respect to both state and Commonwealth legislation, as highlighted in Part II of this article, few cases have been successful. In most cases, courts have concluded that the legislation is not inconsistent with the *RDA*. This is, in part, due to a failure to give the *RDA* a broad, liberal and purposive interpretation. The broader implication of these failures may be that ‘constitutional’ statutes are only as useful as the interpretation of their provisions permits. It is this interpretive effect, in bringing about a broad, liberal and purposive interpretation, that the *RDA* appears to be lacking.

As the analysis below will demonstrate, though judges extra-curially have emphasised the ‘special’ nature of the *RDA*, this ‘special’ status has had little effect in giving effective meaning to the *RDA*’s key terms. In particular, there is a lack of clarity surrounding the meaning of where a law disadvantages or treats persons unequally on the basis of race.¹⁴⁶ In the context of the interpretation of discrimination legislation in Australia more generally, the High Court has accepted that in construing human rights legislation, the courts have a special responsibility to take account of, and give effect to, the statutory purpose.¹⁴⁷ This has been articulated as a ‘fair, large and liberal’ interpretation.¹⁴⁸ This articulation seems similar to the broad, liberal and purposive approach referred to above in Part III(A). The problem in the interpretation of s 10 is not that it is necessarily narrow, but that though the interpretation is broad, this breadth comes at the expense of any depth of substance

¹⁴² *Attorney General v National Assembly for Wales Commission* (n 141) 815 [80] (Lord Hope DPSC).

¹⁴³ Feldman, ‘Statutory Interpretation and Constitutional Legislation’ (n 62) 496–7.

¹⁴⁴ *Ibid.*

¹⁴⁵ *Ibid.*

¹⁴⁶ *Gerhardy* (n 70) 83 (Gibbs CJ).

¹⁴⁷ *Waters v Public Transport Corporation* (1991) 173 CLR 349, 372 (Brennan J), 394 (Dawson and Toohey JJ) and 407 (McHugh J).

¹⁴⁸ D Pearce, *Statutory Interpretation in Australia* (LexisNexis Butterworths, 9th ed, 2014) 314–15, quoting *IW v City of Perth* (1997) 191 CLR 1, 11 (Brennan CJ and McHugh J).

or purpose to its terms. The High Court first examined s 10 of the *RDA* in *Gerhardy*. In that case, the Court considered whether the *Pitjantjatjara Land Rights Act 1981* (SA) was racially discriminatory where it made it an offence for a person who was not a Pitjantjatjara to enter land without the permission of Anangu Pitjantjatjara.¹⁴⁹ The Court concluded that the Act did create a distinction based on race,¹⁵⁰ but was nevertheless 'saved' as a special measure.¹⁵¹ Sadurski criticised the judgments in *Gerhardy* for failing to appreciate the differences between a distinction based on race and racial discrimination.¹⁵² He argued the judgments in *Gerhardy*, for the most part, only engaged with the concept of discrimination in a cursory or simplistic manner.¹⁵³ This is a failure that has continued, and been exacerbated, since that time. It is in the failure to determine the link between a distinction based on race and discrimination and equality that demonstrates the problems of an interpretation that is neither liberal nor purposive.

A Law which Disadvantages Persons based on Race

A key question when determining whether a law is inconsistent with s 10 is whether it creates a 'distinction on the basis of race' or whether a law disadvantages or creates inequality on the basis of race.¹⁵⁴ In interpreting s 10, the High Court has adopted an interpretation marked by a high level of generality. Since the early 2000s, the case law has demonstrated a reluctance to link the concept of a distinction or disadvantage to the twin notion of discrimination or broader values such as equality.¹⁵⁵ The consequence of this generality is that there is a lack of analysis or exploration of what kinds of behaviours, practices and laws constitute a distinction on the basis of race, especially when such a distinction involves an intermingling of various aspects of disadvantage.

To constitute a distinction or disadvantage pursuant to s 10 of the *RDA*, a complainant must demonstrate that a law distinguishes on the basis of race or ethnic origin. In determining what constitutes a distinction, or creates a disadvantage, or treats persons unequally on the basis of race the High Court in *Gerhardy*,¹⁵⁶ *Mabo (No 1)*,¹⁵⁷ *Ward*,¹⁵⁸ and *Maloney*¹⁵⁹ accepted that laws that conflict with the right contained in s 10 are not simply laws that have a purpose of nullifying a person's rights and freedoms on the basis of race, but also include laws that have the effect of distinguishing on the basis of race.

¹⁴⁹ *Gerhardy* (n 70) 75–8 (Gibbs CJ).

¹⁵⁰ *Ibid* 82–3 (Gibbs CJ), 100–1 (Mason J), 122–3 (Brennan J).

¹⁵¹ *Ibid* 86–9 (Gibbs CJ), 103–6 (Mason J), 106–8 (Murphy J), 108–14 (Wilson J), 114–43 (Brennan J), 143–54 (Deane J) and 154–62 (Dawson J).

¹⁵² Wojciech Sadurski, 'Gerhardy v Brown v the Concept of Discrimination: Reflections on the Landmark Case that Wasn't' (1986) 11(1) *Sydney Law Review* 5, 27.

¹⁵³ *Ibid* 28–9.

¹⁵⁴ *Ward* (n 71) 103 (Gleeson CJ, Gaudron, Gummow and Hayne JJ).

¹⁵⁵ See *Ward* (n 71) 99 (Gleeson CJ, Gaudron, Gummow and Hayne JJ); *Maloney* (n 100) 200 (Hayne J).

¹⁵⁶ *Gerhardy* (n 70) 99 (Mason J).

¹⁵⁷ *Mabo (No 1)* (n 113) 230 (Deane J).

¹⁵⁸ *Ward* (n 71) 99 (Gleeson CJ, Gaudron, Gummow and Hayne JJ).

¹⁵⁹ *Maloney* (n 100) 179 (French CJ), 206 (Hayne J), 244 (Bell J), 284 (Gageler J).

The acceptance that s 10 is focused on the effect, rather than the purpose, of the law in question has meant that s 10 can target legislation that creates distinctions *in effect*, as well as explicit distinctions in the text of the legislation on the basis of race.¹⁶⁰ For example, the High Court decision of *Maloney* considered whether liquor restrictions on Palm Island were racially discriminatory though the regulations did not explicitly target Indigenous people. Five justices of the High Court accepted that though the impugned provisions of the *Liquor Act 1992* (Qld) and its associated regulations did not discriminate on their face, the effect of such regulations on the community of Palm Island, where the residents were overwhelmingly Indigenous, had the effect of impairing a person's fundamental rights and freedoms on the basis of race.¹⁶¹

What is noticeable in the judgments on the right contained in s 10 is that the High Court appears resistant to the interpretation or labelling of s 10 as concerned with discrimination on the basis of race.¹⁶² In *Ward*, a case concerning the extinguishment of native title, the majority accepted that s 10 was not concerned with *discrimination* per se, but any distinctions made on the basis of race that could impair fundamental rights and freedoms.¹⁶³ In coming to this conclusion, the majority in *Ward* emphasised that s 10 does not use the word 'discriminatory' or any cognate expressions.¹⁶⁴ Later in the same judgment, the majority distinguished s 10 from other Australian anti-discrimination laws, emphasising that unlike other anti-discrimination law, race is not an irrelevant characteristic for the purpose of the *RDA*, but is something that is required to be considered in determining the purpose or effect of the law in question.¹⁶⁵

These themes are present in Hayne J's judgment in *Maloney*. In *Maloney*, Hayne J emphasised that the subject of s 10 is not 'discrimination',¹⁶⁶ though his Honour noted that the term 'discrimination' is utilised throughout the authorities in which s 10 is discussed.¹⁶⁷ As becomes apparent later in Hayne J's judgment, his Honour's concern with the utilisation of terminology associated with discrimination and anti-discrimination law was that such association would 'inadvertently narrow or confine the operation of s 10' of the *RDA*.¹⁶⁸ This concern appears to be based on an understanding of discrimination law as only being applicable to distinctions that are disproportional or unjustifiable, leaving untouched a range of laws that nevertheless distinguish on the basis of race.¹⁶⁹

¹⁶⁰ Beth Gaze, 'What Vision of Equality and Racial Discrimination Law Does *Maloney v R* Reveal?' (Paper presented at the Roundtable Making Sense of *Maloney v R* (2013), Melbourne Law School, The University of Melbourne, 31 October 2014).

¹⁶¹ *Maloney* (n 100) 191 (French CJ), 206 (Hayne J), 243 (Bell J) and 302 (Gageler J). On this issue Crennan J agreed with Hayne J at 213.

¹⁶² *Ward* (n 71) 99 (Gleeson CJ, Gaudron, Gummow and Hayne JJ).

¹⁶³ *Ibid.*

¹⁶⁴ *Ibid.*

¹⁶⁵ *Ibid* 105 (Gleeson, Gaudron, Gummow and Hayne JJ).

¹⁶⁶ *Maloney* (n 100) 200–2 (Hayne J). For discussion, see Rachel Gear, 'Commentary: Alcohol Restrictions and Indigenous Australians: The Social and Policy Implications of *Maloney v The Queen*' (2014–2015) 21(1) *James Cook University Law Review* 41, 47.

¹⁶⁷ *Maloney* (n 100) 200. Though, noticeably there is no authority cited for this statement.

¹⁶⁸ *Ibid* 201–2.

¹⁶⁹ *Ibid.*

These concerns, as to the possible limitations of the concept of discrimination, are not unfounded. There are numerous critiques that other anti-discrimination legislation's complex, artificial and obscure language has unnecessarily stymied the development of non-discrimination principles in Australia.¹⁷⁰ But, the approach to s 10 illustrates the opposite problem. While its broad terms allow for a variety of different kinds of legislation to potentially fall foul of s 10 where there may be a difference in effect based on race, there is no indication in the case law as to what constitutes a distinction *in effect*. Courts have accepted that laws that entirely exclude persons from exercising a right or freedom, or laws that absolutely extinguish a right or freedom for persons based on race are captured by the terms of s 10.¹⁷¹ In *Maloney*, a majority of justices were willing to accept the liquor regulations created a distinction where the vast majority of the persons affected were Indigenous due to the geographical location of the order.¹⁷² However, where the distinction occurs due to race or ethnic origins *combined* with stereotypes and other facets of socio-economic disadvantage predicated on prolonged and systemic issues of racial injustice, the case law struggles to apply the broad principles of distinction to capture the effect of such laws.

In *Aurukun Shire Council v Chief Executive Officer, Office of Liquor Gaming and Racing in the Department of Treasury*, Keane JA rejected an argument that s 10 could be applied in a manner that was cognisant and reflective of economic and geographical circumstances.¹⁷³ *Aurukun* involved the refusal to issue liquor licenses to local councils in circumstances in which the appellant councils operated the only taverns in their respective local areas.¹⁷⁴ Keane JA concluded that while s 10 was concerned with the practical effect of the law, rather than formal expression, the practical effect of the impugned provision was that no resident in Queensland was able to acquire alcohol from their local government. What the appellants were complaining of, according to Keane JA, was not a distinction based on 'race', but a 'consequence of the different geographical and socio-economic conditions which obtain, and which have obtained for many years, in different areas of the State'.¹⁷⁵ As such, his Honour concluded that the purpose of s 10 was not to remedy the 'serious level of relative socio-economic disadvantage which affects the appellants' communities'.¹⁷⁶

A similar failure to consider in any detail the interplay between racial discrimination and socio-economic disadvantage is also present in the decisions from the NT Supreme Court in *R v Woods*¹⁷⁷ and Court of Appeal in *Munkara v Bencsevich*.¹⁷⁸ In *Woods*, the Full Court of the NT Supreme Court considered whether the *Juries Act 1962* (NT) was inconsistent with ss 9 or 10 of the *RDA*. The

¹⁷⁰ See, eg, the critiques made in *IW v City of Perth* (n 148) 12 (Brennan CJ and McHugh J), 37 (Gummow J).

¹⁷¹ See *Mabo (No 1)* (n 113).

¹⁷² *Maloney* (n 100) 191 (French CJ) 206 (Hayne J, Crennan J agreeing at 213), 244 (Bell J), 302 (Gageler J).

¹⁷³ *Aurukun* (n 121) 73–4 (Keane JA).

¹⁷⁴ *Ibid* 4–5, 7.

¹⁷⁵ *Ibid* 75 (Keane JA).

¹⁷⁶ *Ibid* 76 (Keane JA).

¹⁷⁷ *R v Woods* (2010) 246 FLR 4 ('*Woods*').

¹⁷⁸ *Munkara v Bencsevich* [2018] NTCA 4 ('*Munkara*').

appellants argued that the Act was inconsistent with the *RDA* on the basis that it disqualified from jury service persons in custody within the previous seven years.¹⁷⁹ The appellants argued that this disqualification disproportionately affected Indigenous Australians.¹⁸⁰ As 83% of the NT prison population was Indigenous, the appellants argued that such a preclusion from jury service created a distinction based on race.¹⁸¹ The Full Court of the NT Supreme Court rejected the appellants' arguments on the basis that such disqualification of Indigenous jurors would, in any event, not impair the appellants' right to a fair trial.¹⁸² However, the Court also rejected the argument that there was a distinction based on race in any event.¹⁸³ This was on the basis that there was no direct evidence as to how many persons in any single case would be captured by such a preclusion, and evidence based on statistics and 'usual experience' did not support such an inference.¹⁸⁴ As such, a 'distinction' based on race could not be supported by the evidence.¹⁸⁵

In *Munkara*, the appellant challenged provisions of the *Alcohol Protection Orders Act 2013* (NT) on the basis that the practical operation of such provisions meant that Indigenous persons did not enjoy the right to freedom of movement, access to public places, privacy and equal treatment to the same extent as persons of other races.¹⁸⁶ The Act allowed police to issue alcohol protection orders to persons who had committed offences while affected by alcohol.¹⁸⁷ The appellant argued that the impugned provisions created a practical distinction on the basis of race.¹⁸⁸ This was on the basis that, as accepted at first instance, the persons who came 'within the net' of the Act were overwhelmingly Indigenous.¹⁸⁹ It was accepted that Indigenous Northern Territorians were 'overwhelmingly more likely to be arrested, summonsed or served with a notice to appear in court in respect of a qualifying offence'.¹⁹⁰ The primary judge accepted that 86% of the protection orders issued had been issued to Indigenous persons.¹⁹¹ However, the primary judge and the NT Court of Appeal rejected that the statistical evidence of the practical effect of the law demonstrated a distinction on the basis of race.¹⁹² The Court of Appeal concluded that the distinction in effect was not based on race, but on the consequences of behaviour (the committing of a qualifying offence).¹⁹³ The Court of Appeal labelled the arguments made by the appellants regarding the interplay between race, disadvantage, alcohol and interaction with the police as 'simplistic' and 'offensive'.¹⁹⁴

¹⁷⁹ *Woods* (n 177) 10.

¹⁸⁰ *Ibid.*

¹⁸¹ *Ibid* 9.

¹⁸² *Ibid* 20.

¹⁸³ *Ibid* 19.

¹⁸⁴ *Ibid* 19–20.

¹⁸⁵ *Ibid* 20.

¹⁸⁶ *Munkara* (n 178) [9] (Kelly J).

¹⁸⁷ *Ibid* [3] (Kelly J).

¹⁸⁸ *Ibid* [12] (Kelly J).

¹⁸⁹ *Ibid* [95]–[96] (Blokland J).

¹⁹⁰ *Ibid.*

¹⁹¹ *Ibid.*

¹⁹² *Ibid* [16] (Kelly J), [96], [99] (Blokland J).

¹⁹³ *Ibid* [99]–[101] (Blokland J).

¹⁹⁴ *Ibid* [102] (Blokland J).

These decisions give rise to two interrelated problems: one evidentiary and one conceptual. The evidentiary problem is that though judicial decisions relating to s 10 acknowledge that the law is concerned with the practical effect of the challenged provisions, there is little guidance provided as to what kind of evidence a claimant can provide to demonstrate a distinction *in effect* that is singly concerned with race. From *Woods* and *Munkara*, it appears that evidence of a statistical disparity in effect is not sufficient to demonstrate a practical distinction based on race.¹⁹⁵ Nor, drawing on the reasoning of the majority of *Aurukun*, does a distinction exist where it is also justified and explained on the basis socio-economic disadvantage.¹⁹⁶ Instead, these decisions seem to indicate that where a provision applies equally to all persons across a jurisdiction, the distinction must be related to something ‘directly’ related to race without a consideration of disadvantage to which historical and systemic racial injustice has contributed.¹⁹⁷ The ultimate outcome of such an approach to evidence is that the interpretation of s 10 to include effect is impotent given that effect will invariably have several causes many of which relate indirectly to race.¹⁹⁸ This evidentiary problem is linked to a broader conceptual issue with the constructed tests for demonstrating a breach of s 10. The test developed to determine inconsistency with s 10 of the *RDA* is so broad that it is difficult to identify what distinctions are captured by s 10 and why such distinctions are wrongful and should be prohibited.

IV The Interpretation of the *Racial Discrimination Act 1975* (Cth) within the Broader Constitutional Culture

The criticism that the interpretation of s 10 makes it ineffectual in preventing or limiting discrimination is consistent with academic criticisms about anti-discrimination law generally in Australia. Much of the academic commentary on the interpretation of discrimination law by the judiciary, in Australia and elsewhere, is often critical and focuses on the courts’ failure to give the terms of discrimination law a broad and substantive meaning.¹⁹⁹ However, the interpretation of s 10 is different from the problems associated with other aspects of anti-discrimination law in Australia. In this section I will articulate those differences and place the interpretation of s 10 within Australia’s broader constitutional culture and interpretation. I will demonstrate that though the interpretation of s 10 is unusual when compared with its foreign equivalents and other aspects of Australian

¹⁹⁵ *Woods* (n 177) 19 (Full Court); *Munkara* (n 178) [102] (Blokland J).

¹⁹⁶ *Aurukun* (n 121) 76 (Keane JA). See also Fiona Campbell, ‘Deficit Discourse — The ‘Regime of Truth’ preceding the Cape York Welfare Reform’ (2019) 28(3) *Griffith Law Review* 303, 313.

¹⁹⁷ *Munkara* (n 178) [96] (Blokland J); *Woods* (n 177) 19 (The Court). For a discussion of the interpretation of discrimination law with an eye to systemic and historical injustice, see: Margaret Thornton, *The Liberal Promise: Anti-Discrimination Legislation in Australia* (Oxford University Press, 1990) 44–6 (‘*The Liberal Promise*’); Sandra Fredman, *Discrimination Law* (Oxford University Press, 2nd ed, 2011) 38.

¹⁹⁸ *Maloney* (n 100) 206 (Hayne J).

¹⁹⁹ See, eg, Thornton, *The Liberal Promise* (n 197) 212–3; Beth Gaze, ‘Context and Interpretation in Anti-Discrimination Law’ (2002) 26(2) *Melbourne University Law Review* 325 (‘Context and Interpretation in Anti-Discrimination Law’); Rees, Rice and Allen (n 72) 26–7.

anti-discrimination law, it is consistent with the broader approach to values in Australian constitutional law.

Reasons attributed to the judiciary's failure to develop a more substantive account of discrimination law include the prescriptive legislative text,²⁰⁰ and the joint failure of the legislature and the judiciary to develop a clear account of the aims and purposes of discrimination law.²⁰¹ The prescriptive nature of the legislative text has been raised in both the commentary and by the High Court as a rationale for the limited manner in which anti-discrimination statutes have been interpreted. In *IW v City of Perth*, Brennan CJ and McHugh J acknowledged the need to give the legislation a broad interpretation pursuant to both general rules of statutory interpretation and the requirements of the *Acts Interpretation Act 1984* (WA), but cautioned that

[g]iven the artificial definitions of discrimination in the Act and the restricted scope of their application, the court or tribunal should not approach the task of construction with a presumption that conduct which is discriminatory in its ordinary meaning is prohibited by the Act. The Act is not a comprehensive anti-discrimination or equal opportunity statute. The legislature of Western Australia, like other legislatures in Australia and the United Kingdom has avoided use of general definitions of discrimination.²⁰²

Considering the interpretation of other discrimination Acts such as the *Sex Discrimination Act 1984* (Cth), Smith suggests a reason that a more substantive or progressive interpretation of discrimination has not emerged is because of the prescriptive and ultimately restrictive nature of the legislative text.²⁰³ This contrasts with the position in Canada, in which the more 'open' nature of the statutory text grants the judiciary a more active and possibly creative role in the construction of terms such as 'discrimination' or 'distinction'. For example, the *Canadian Human Rights Act* does not create a distinction between 'direct' and 'indirect' discrimination in the manner of most Australian discrimination law statutes.²⁰⁴ Instead, the *Canadian Human Rights Act* outlines the characteristics protected under the Act and prohibits discrimination in a variety of areas of public life, but the term 'discrimination' itself is undefined.²⁰⁵ Smith argues that it is this more open language of the Canadian Acts that allows the judiciary a greater role in the interpretation and development of discrimination law.²⁰⁶ While this argument may hold for the other Commonwealth Acts and for other, more prescriptive provisions of the *RDA*, it does not appear to explain the judicial approach with respect to the interpretation of s 10. In contrast to other provisions in Australian anti-discrimination law, s 10 is not prescriptive and its terms are wide and general in nature. Consequently, there is significantly more scope to interpret it in a broad, substantive manner with an eye to

²⁰⁰ Belinda Smith, 'Rethinking the Sex Discrimination Act' in Margaret Thornton (ed) *Sex Discrimination in Uncertain Times* (ANU Press, 2010) 235, 235–6; Rees, Rice and Allen (n 72) 52–3, 88–9.

²⁰¹ Gaze, 'Context and Interpretation in Anti-Discrimination Law' (n 199) 331–2; Rees, Rice and Allen (n 72) 7.

²⁰² *IW v City of Perth* (n 148) 14.

²⁰³ Smith (n 200) 238–9.

²⁰⁴ *Canadian Human Rights Act*, RSC 1985, c H-6, s 3.

²⁰⁵ *Ibid.*

²⁰⁶ Smith (n 200) 238.

the international conventions that it implements. Nevertheless, a substantive and clear jurisprudence has not been developed.

Another rationale for the limited manner in which discrimination law is often interpreted is due to the notion of a conservative judicial culture.²⁰⁷ This culture makes judges reluctant to interpret progressive legislation in a way that extends and expands on its progressive aims.²⁰⁸ To better understand the impact of that conservative judicial culture and its association with a broader constitutional culture, it may again be useful to contrast the interpretation of the *RDA* and its perceived specialness with the interpretation of constitutional statutes in comparable jurisdictions. As highlighted in Part I of this article, one justification for determining that a statute is constitutional is that it is fundamental law and draws on constitutional values.²⁰⁹ This justification has been particularly pertinent in the context of the Canadian human rights codes. In the interpretation of the codes, the Supreme Court of Canada invokes and expands on the constitutional values, in particular, the constitutional value of equality.²¹⁰ In the interpretation of s 10, while the ideas of a law providing distinction or disadvantage on the basis of race which impairs fundamental rights and freedoms is interpreted with a high degree of breadth and generality — consistently with constitutional or quasi-constitutional statutes in other jurisdictions,²¹¹ this interpretation is not with an eye to the underlying fundamental aspirations or values behind it. From the s 10 jurisprudence of both the High Court and state and territory courts of appeal, there is a lack of clarity as to the reasons why race-based distinctions are problematic and as to the way in which such distinctions unfairly impact a person's life. It is in the interpretation of the rights contained in the *RDA* where there is a divergence in approach to 'constitutional' statutes. While in both Canada and the UK, 'constitutional' statutes are interpreted with an eye to fundamental constitutional values, this is not apparent in the interpretation of the *RDA*. While the interpretation of s 10 is broad in scope, it lacks a clear or coherent articulation of the underlying purpose of the prohibition contained within it. This lack of coherence is, however, consistent with the interpretive approach taken to values in the Australian constitutional context.

While the *Australian Constitution* does play a role in defining and preserving fundamental values, the role it plays is often presented as 'thin' or 'muted'.²¹² Though, over time, the High Court's constitutional jurisdiction has shifted focus and does, to a degree, articulate and enforce fundamental values underlying the *Australian Constitution*, the role those values play is still limited. As Dixon has highlighted, even in the consideration of explicit rights in the *Australian Constitution* such as the right to trial by jury provided by s 80 or freedom of religion in s 116, the High Court has taken a narrow approach to their interpretation without a focus on

²⁰⁷ Margaret Thornton, 'Disabling Discrimination Legislation: The High Court and Judicial Activism' (2009) 15(1) *Australian Journal of Human Rights* 1, 3 ('Disabling Discrimination Legislation'); Gaze, 'Context and Interpretation in Anti-Discrimination Law' (n 199) 341.

²⁰⁸ Thornton, 'Disabling Discrimination Legislation' (n 207) 21; Gaze, 'Context and Interpretation in Anti-Discrimination Law' (n 199) 341.

²⁰⁹ See above Part II(A).

²¹⁰ Helis (n 7) 177, 179–80.

²¹¹ Ibid 17.

²¹² Elisa Arcioni and Adrienne Stone, 'The Small Brown Bird: Values and Aspirations in the *Australian Constitution*' (2016) 14(1) *International Journal of Constitutional Law* 60, 60.

their underlying ‘fundamental’ nature.²¹³ In addition, though there is now a greater focus on fundamental values in the *Australian Constitution*, as Roux has commented, this focus has not necessarily led to a greater candour in the legal reasoning process:

To the extent that those reforms were aimed at introducing greater candor about the role of extralegal values in the judicial reasoning process, they failed. In times of trouble, High Court justices’ instinct is still to fall back on a conception of law as a technically exacting discipline capable of generating political neutral answers to controversial questions. To that extent, a version of democratic legalism premised on the denial of law’s politicality still holds sway.²¹⁴

This thinness and lack of transparency is apparent when the High Court considers the twin notions of equality and discrimination in the constitutional context. As Simpson has commented with respect to the term ‘discrimination’ in a constitutional setting, the High Court’s jurisprudence is marked by a high degree of generality, and the concept is defined in ‘universal’ and ‘abstract’ terms.²¹⁵ In that context, such an approach leads to a preference for a test to determine discrimination that provides inadequate guidance when faced with the reality and contours of a non-discrimination question.²¹⁶ More broadly, the general approach to questions of values in Australian constitutional law is to view them as a skeleton without an interrogation of their underlying meaning and application.

Utilising values as a skeleton has meant that while the High Court has drawn on overarching values such as equality and non-discrimination to justify conclusions, this has not led to a significant discussion of what such values entail. For example, the value of non-discrimination was utilised in *Mabo v Queensland (No 2)*.²¹⁷ Brennan J, in particular, concluded that it was imperative to ensure that the common law was not ‘frozen in an age of racial discrimination’.²¹⁸ His Honour argued that the courts were giving effect to ‘the enduring community value of non-discrimination, that is, the equality of all people before the law’.²¹⁹ This value of non-discrimination was described as ‘the skeleton of principle, which gives the body of our law its shape and internal consistency’.²²⁰ At times, there has been an attempt to give the values of equality and discrimination more substantive depth, such as in Gaudron J’s judgment in *Street v Queensland Bar Association*,²²¹ or in the dissenting judgments of Deane and Toohey JJ in *Leeth v Commonwealth*.²²² Nevertheless, while equality is still a value underlying Australian constitutionalism,

²¹³ Rosalind Dixon, ‘An Australian (Partial) Bill of Rights’ (2016) 14(1) *International Journal of Constitutional Law* 80, 97.

²¹⁴ Theunis Roux, *The Politico-Legal Dynamics of Judicial Review: A Comparative Analysis* (Cambridge University Press, 2018) 133.

²¹⁵ Amelia Simpson, ‘The High Court’s Conception of Discrimination: Origins, Applications and Implications’ (2007) 29(2) *Sydney Law Review* 263, 288.

²¹⁶ *Ibid.*

²¹⁷ *Mabo v Queensland (No 2)* (1992) 175 CLR 1 (*‘Mabo (No 2)’*).

²¹⁸ *Ibid* 42 (Brennan J).

²¹⁹ *Ibid.*

²²⁰ *Ibid* 29.

²²¹ *Street v Queensland Bar Association* (1989) 168 CLR 461, 566 (Gaudron J).

²²² *Leeth v Commonwealth* (1992) 174 CLR 455, 486–7 (Deane and Toohey JJ) (*‘Leeth’*).

it remains a skeleton principle without the content necessary for a substantive interpretation of statutory discrimination law.

Brennan J utilised these values to justify the conclusion in *Mabo (No 2)*.²²³ But, in *Dietrich v The Queen*, determined less than five months later, his Honour warned of the use of ‘contemporary values’ in justifying judicial development:

The contemporary values which justify judicial development of the law are not the transient notions which emerge in reaction to a particular event or which are inspired by a publicity campaign conducted by an interest group. They are the relatively permanent values of the Australian community. Even if the perception of contemporary values is coloured by the opinions of individual judges, judicial experience in the practical application of legal principles and the coincidence of judicial opinions in appellate courts provide some assurance that those values are correctly perceived. The responsibility for keeping the common law consonant with contemporary values does not mean that the courts have a general power to mould society and its institutions according to judicial perceptions of what is conducive to the attainment of those values.²²⁴

This resistance to ‘mould[ing] society and institutions according to judicial perceptions of what is conducive to the attainment of those values’²²⁵ has come at the expense of articulating the substance of those values. In particular, there is no articulation of what these values require of other branches of government and private parties. Instead, the values of equality and non-discrimination are referenced in the abstract to justify certain conclusions without any engagement with their content.

While the values of equality and non-discrimination are utilised to justify a particular conclusion, the role of the courts in elaborating, articulating or expanding the scope of these values is limited. In the constitutional context, it is possible to interpret the High Court’s jurisprudence as demonstrating a wide variation of values on a continuum.²²⁶ The continuum extends from formal to substantive non-discrimination between individuals, as well as demarcating a divide between a formal and substantive commitment to the equality of citizens within the federal compact.²²⁷ But what is understood as substantive in the constitutional context is similar to the approach adopted pursuant to s 10 of the *RDA* outlined above, in that the High Court considers the *effect* of a distinction or a difference, rather than only its legal form.²²⁸ This ‘substantive’ approach nevertheless still gives little articulation of the underlying rationale for why such distinctions are problematic, nor does the case law provide a clear test or guidance in determining substantive facets or factors of discrimination.

In the main, the High Court’s approach to discrimination and equality in the context of constitutional values is simply to acknowledge their existence and the role they play in legal reasoning. But the Court still fails to give these values any depth

²²³ *Mabo (No 2)* (n 217) 30 (Brennan J).

²²⁴ *Dietrich v The Queen* (1992) 177 CLR 292, 319 (Brennan J).

²²⁵ *Ibid.*

²²⁶ Amelia Simpson, ‘Equal Treatment and Non-Discrimination’ in Rosalind Dixon (ed) *Australian Constitutional Values* (Hart Publishing, 2018) 195, 197.

²²⁷ *Ibid.*

²²⁸ *Ibid.* 200.

or meaning. Particularly with respect to the twin values of non-discrimination and equality, even at its most radical, the jurisprudence concludes the possible existence of legal equality in Australian constitutional law without significant interrogation as to its meaning or application.²²⁹ The Australian approach to overarching values as a kind of skeleton structure means that the articulation of what constitute equality fails to answer any of the key questions that the *RDA* grapples with as to its substance and conceptual underpinnings. Instead, equality is simply associated with a notion of fairness.²³⁰

The interpretation of s 10 of the *RDA* fails to identify key indicia of discrimination because the jurisprudence on s 10 fails to grapple with the values underlying the *RDA* such as discrimination and equality. Understanding the limitations of the interpretation and the substantive effect of the *RDA* is important to assess the utility in strengthening the ‘constitutional’ force of the *RDA* as has been suggested by Lino.²³¹ Ultimately, it is not the lack of ‘constitutional’ force that has led the *RDA* to have a more limited effect than the rights-protecting documents of comparable jurisdictions, but its interpretation. That interpretation is consistent with the interpretation of values more broadly in Australian law. As a consequence, any attempt to strengthen the force of the *RDA* through providing for manner and form provisions is unlikely to change the underlying problems with its interpretation.

V Conclusion

At the outset of this article, I highlighted that since its passage, the *RDA* had been considered a special piece of legislation. Its specialness stems from the idea that it provides a form of rights protection by prohibiting race discrimination and from its effect on the lawmaking powers of state legislatures. It is these factors, combined with the *RDA*’s stickiness in public culture that gives it an almost constitutional force. While the *RDA* does demonstrate that the legislative process can be harnessed to achieve a degree of recognition of the importance of non-discrimination and equality, the interpretation of the *RDA* equally demonstrates its limitations. While it is understandable to hope that the *RDA* is reflective of an underlying commitment to non-discrimination and equality, to be effective in actually achieving a commitment to racial equality, there needs to be a more rigorous assessment of the effectiveness of the legislative tools currently in use. In order to be effective, those tools must be interpreted with an eye to systemic and historical issues of disadvantage,²³² and a stricter interrogation of executive and legislative action.

In this article, I interrogated whether the *RDA* could be considered a special or ‘constitutional’ form of legislation. In the public consciousness and by the legislature, the *RDA* has been treated as a form of special legislation. It has been difficult to amend to limit its terms, and Commonwealth legislatures have explicitly

²²⁹ *Leath* (n 222) 486–7 (Deane and Toohey JJ). The judgment was described as ‘radical’ by Kirk: Jeremy Kirk, ‘Constitutional Implications (II): Doctrines of Equality and Democracy’ (2001) 25(1) *Melbourne University Law Review* 24, 24.

²³⁰ Chief Justice Allsop, ‘Values in Public Law’ (2017) 91(2) *Australian Law Journal* 118, 119–20.

²³¹ Lino, ‘Thinking outside the *Constitution*’ (n 4) 384.

²³² For discussion, see Fredman (n 197) 26–7.

excluded its operation where other legislation may fall foul of its provisions. Based on a range of definitions, focused on its purpose, form, function and public consciousness, I concluded that the *RDA* could be considered a form of constitutional legislation. This, in turn, has been reflected, at times, in its interaction with other statutes at both a state and federal level, with courts applying limitations on the principle of implied repeal and a derogation from the general principle of *generalia specialibus non derogant* where there has been inconsistency between the *RDA* and other statutes.

However, though this could provide the *RDA* with a strong protective quality, this is stymied by the manner in which the terms of s 10 have been interpreted. Such that few other pieces of legislation have been found to be inconsistent with s 10. While the *RDA* has been given a broad interpretation on its terms, this is nevertheless not necessarily to the benefit of vulnerable groups based on race and ethnic origin. Instead, the breadth in interpretation has led to a failure to interrogate the nature of equality and its capacity to challenge and ameliorate underlying systemic racial disadvantage.

Though I do not deny the symbolic importance of legislation such as the *RDA*, unless this symbolism leads to a purposive interpretation of the legislative text, there appears little value in its classification as special or constitutional in the Australian context. There is a danger in the important symbolic effect of the legislation not being reflected in its legal interpretation.²³³ The danger is that while the public and legislatures may believe that the *RDA* has significant force, without a change in interpretation the law will continue to be ineffective in creating broader and more substantive change for claimants in race discrimination claims.

²³³ I thank an anonymous reviewer for emphasising this final point.

Accuracy, Utility and Gateways: Justifications(?) for Controlling the Use of Surrounding Circumstances in Contractual Interpretation

NA Tiverios*

Abstract

The status of the principle of contractual interpretation enunciated by Mason J in *Codelfa Construction Pty Ltd v State Rail Authority of NSW* (1982) 149 CLR 337 is uncertain. That principle being that ambiguity is first required when interpreting a contract before recourse can be had to evidence of the particular ‘surrounding circumstances’ known to the parties at the time of entering into the agreement. In this article, I assess from first principles the desirability of this so-called ‘ambiguity gateway’. I draw on developments in the philosophy of language and mind to illustrate how the ambiguity gateway detracts from the interpretive process. I then consider to what extent the ambiguity gateway is justifiable on the basis of making contractual disputes more efficient *in globo* (that is, on the basis of traditional rule-based utilitarianism). I conclude that it is incumbent on those making this utilitarian claim to justify their conclusion that the ambiguity gateway performs an efficiency enhancing function. Given the sceptical arguments presented in this article, it is doubtful that the ambiguity gateway will ever be justified on the basis of rule-based utilitarianism. In the absence of such a justification, the principle should be abolished.

* Senior Lecturer, Law School, University of Western Australia, Australia; Barrister and Solicitor, Supreme Court of the Australian Capital Territory and High Court of Australia. Email: nicholas.tiverios@uwa.edu.au; ORCID iD: <https://orcid.org/0000-0002-9067-0686>

This article benefited from comments from participants at the University College London New Work in Obligations Conference, presenting at the Faculty of Law, University of New South Wales (‘UNSW’) and from the students in my Advanced Contract Law course at UNSW (2020). I would like to thank the anonymous reviewers and editors for their comments. All errors are my own.

Please cite this article as:

NA Tiverios, 'Accuracy, Utility and Gateways: Justifications(?) for Controlling the Use of Surrounding Circumstances in Contractual Interpretation' (2021) 43(4) *Sydney Law Review* 547.



This work is licensed under a Creative Commons Attribution-NonDerivatives 4.0 International Licence (CC BY-ND 4.0).

As an open access journal, unmodified content is free to use with proper attribution. Please email sydneylawreview@sydney.edu.au for permission and/or queries. © 2021 Sydney Law Review and authors. ISSN: 1444-9528

I Introduction

My purpose in this article is to consider to what extent the principle of contractual interpretation set out by the High Court of Australia in *Codelfa Construction Pty Ltd v State Rail Authority of NSW*¹ is justifiable. That principle being that ambiguity is first required when interpreting a contract before recourse can be had to evidence of the particular 'surrounding circumstances' known to the parties at the time of entering into the agreement. This principle has become known colloquially by Australian lawyers as the 'ambiguity gateway'.² I will adopt this moniker here. The status of the ambiguity gateway remains contentious. A significant amount of ink has been spilt on this issue by academics,³ practitioners,⁴ and within the growing corpus of case law itself.⁵ The principles applicable remain uncertain and they seemingly differ between the various intermediate appellate courts in Australia.

¹ *Codelfa Construction Pty Ltd v State Rail Authority of NSW* (1982) 149 CLR 337, 352 (Mason J, Stephen J agreeing at 344, Wilson J agreeing at 392).

² See, eg, Justice Kenneth Martin, 'Contractual Construction: Surrounding Circumstances and the Ambiguity Gateway' (2013) 37(2) *Australian Bar Review* 118; Justice Robert McDougall, 'Construction of Contracts: The High Court's Approach' (2016) 41(2) *Australian Bar Review* 103.

³ See, eg, David McLauchlan, 'Contractual Interpretation: What Is It About?' (2009) 31(1) *Sydney Law Review* 5 ('Contractual Interpretation: What Is It About?'); David McLauchlan, 'Plain Meaning and Commercial Construction: Has Australia Adopted the ICS Principles?' (2009) 25(1) *Journal of Contract Law* 7; John Carter, 'Context and Literalism in Construction' (2014) 31(2) *Journal of Contract Law* 100; Robert Stevens, 'The Meaning of Words and the Intentions of Persons' in James Edelman, Simone Degeling and James Goudkamp (eds), *Contract in Commercial Law* (Lawbook Co, 2016) 167; John Eldridge, 'Surrounding Circumstances in Contractual Interpretation: Where Are We Now?' (2018) 32(3) *Commercial Law Quarterly* 3.

⁴ See, eg, Martin (n 2); McDougall (n 2); Chief Justice James Spigelman, 'From Text to Context: Contemporary Contractual Interpretation' (2007) 81(5) *Australian Law Journal* 322; Sir Anthony Mason, 'Opening Address' (2009) 25(1) *Journal of Contract Law* 1; Derek Wong and Brent Michael, 'Western Export Services v Jireh International: Ambiguity as the Gateway to Surrounding Circumstances?' (2012) 86(1) *Australian Law Journal* 57; Kevin Lindgren, 'The Ambiguity of Ambiguity in the Construction of Contracts' (2014) 38(2) *Australian Bar Review* 153; Thomas Prince, 'Defending Orthodoxy: *Codelfa* and Ambiguity' (2015) 89(7) *Australian Law Journal* 491 ('Defending Orthodoxy: *Codelfa* and Ambiguity'); Daniel Reynolds, 'Construction of Contracts after *Mount Bruce Mining v Wright Prospecting*' (2016) 90(3) *Australian Law Journal* 190; Thomas Prince, 'Still Defending Orthodoxy: The New Front in the War on *Codelfa*' (2018) 46(1-2) *Australian Bar Review* 156.

⁵ A useful summary of this controversy and a more than sufficient collection of authorities are provided in *Aurizon Network Pty Ltd v Glencore Coal Queensland Pty Ltd* (2019) 1 QR 392, 427-9 [118]-[121] (Jackson J) ('*Aurizon Network*').

The purpose of this article is, however, somewhat distinct from others. It is to set out an underlying framework for assessing whether the ambiguity gateway is justifiable from first principles. This is achieved in two steps. The first step is to identify the most persuasive arguments in favour of maintaining the ambiguity gateway. I argue that the most common argument in favour of the gateway in Anglo-Australian jurisdictions is derived from traditional rule-based utilitarianism: that limiting the prescribed contextual indicators in contractual interpretation cases creates more ‘benefits’ than ‘costs’ by making the resolution of disputes more efficient *in globo* albeit at the expense of contextual interpretive accuracy in *particular* cases. This argument is pitched primarily at a level that transcends a particular interpretive dispute between A and B and considers the legal system as a whole and the consequences of the ambiguity gateway.⁶ On the other hand, those in favour of removing the ambiguity gateway correctly point out that any utterance or communication can have a fundamentally different meaning when isolated from the context in which it was made. As such, those who wish to remove the ambiguity gateway are more concerned with imbuing the contractual rights, duties, powers, liabilities, privileges and immunities between A and B with the greatest level of interpretive accuracy. After all, such jural relations are what the parties to the contract assented to. The important point is that the arguments for retaining and abolishing the ambiguity gateway do not operate at the same conceptual level. It is for this reason that it is all too easy for those starting from such different premises to ‘talk past’ each other when debating this issue.

The second step is to consider to what extent the ambiguity gateway is ‘fit for purpose’. That is, although the gateway has a clear justification, is the rule nonetheless *designed* in a manner that is consistent with its underlying rationale? Given that there is often more than one way that a more abstract and general moral principle can be translated into a directly applicable legal rule, it is possible that the ambiguity gateway is not designed in a manner that properly achieves the efficiency gains sought. Indeed, as I seek to demonstrate in this article, if the purpose of the gateway is to make the resolution of contractual interpretation disputes more efficient, then it can be seriously questioned whether the rule is fit for such a purpose. As such, in this article I do not provide a conclusive view on the ambiguity gateway controversy. Rather, I provide a framework for assessing the desirability of the principle and, in doing so, draw heavily from the experience of other Commonwealth jurisdictions. This is done in three parts. In Part II of this article I set out the current state of the law in Australia. In Part III I set out the normative arguments in favour of both abolishing and maintaining the ambiguity gateway (that is, ‘step one’ and some of ‘step two’ above). Then in Part IV I consider potential alternative approaches to the ambiguity gateway — that is, the rest of step two above: what other approaches could be adopted in Australia in order adequately to balance the competing considerations of interpretive accuracy and economic efficiency.

⁶ This is because an economic analysis of a particular legal issue takes an *ex-ante* perspective. From this perspective desirability of a rule will be evaluated on the basis of the consequences of having the rule for future actors. This stands in contrast to non-consequentialist theories that are concerned with the existence of *a priori* rights.

II Principles – Where is Australia Now?

The principle set out in *Codelfa* is that ambiguity is required when interpreting a contract *before* recourse can be had to extrinsic evidence as to the particular surrounding circumstances known to the parties at the time of entry into the agreement (being the ‘ambiguity gateway’ introduced above).⁷ These surrounding circumstances are understood here as those objective facts that: (i) were reasonably known to both contracting parties at the time the contract was entered into; and (ii) provide relevant evidence of the background and context against which the parties formed the contract. Such evidence could feasibly be used to assist in the interpretative processes of identifying a meaning of a descriptive term, explaining the purpose of the transaction,⁸ or otherwise shedding light on the most likely meaning of an otherwise ambiguous term.⁹ As Mason J (with whom Stephen and Wilson JJ agreed) said in *Codelfa*:

The true rule is that evidence of surrounding circumstances is admissible to assist in the interpretation of the contract if the language is *ambiguous or susceptible of more than one meaning*. But it is not admissible to contradict the language of the contract when it has a plain meaning. Generally speaking facts existing when the contract was made will not be receivable as part of the surrounding circumstances as an aid to construction, unless they were known to both parties, although, as we have seen, *if the facts are notorious knowledge of them will be presumed*.¹⁰

The approach espoused by Mason J has since been reaffirmed by the High Court of Australia on several occasions,¹¹ most notably in *Western Export Services Inc v Jireh International Pty Ltd*.¹² In *Jireh*, the Special Leave Panel of Gummow, Heydon and Bell JJ took the unusual step, when refusing special leave to appeal, to state that:

Acceptance of the applicant’s submission, clearly would require reconsideration by this Court of what was said in *Codelfa Construction Pty*

⁷ For what constitutes ambiguity see below n 95 and accompanying text. The position I take here is that *Codelfa* (n 1) should not be understood as an extension of the parol evidence rule. Rather, the parol evidence rule is concerned with what documents constitutes the parties’ agreement and *not* what evidence could be relevant and probative in discerning the meaning of that agreement. In this connection, compare: Nick Seddon and Rick Bigwood, *Cheshire and Fifoot Law of Contract* (LexisNexis, 11th ed, 2017) 424–5 [10.4] with Edwin Peel and Guenter Treitel, *The Law of Contract* (Sweet & Maxwell, 14th ed, 2015) 233–4 [6-014].

⁸ Such evidence may include the genesis of the transaction, the background, the context and the market in which the parties are operating: *International Air Transport Association v Ansett Australia Holdings Ltd* (2008) 234 CLR 151, 160 [8] (Gleeson CJ) (*‘International Air Transport Association’*).

⁹ See, eg, *DTR Nominees Pty Ltd v Mona Homes Pty Ltd* (1978) 138 CLR 423, 429 (Stephen, Mason and Jacobs JJ). See also McDougall (n 2) 105.

¹⁰ *Codelfa* (n 1) 352 (emphasis added). For a historical overview of this approach see Prince, ‘Defending Orthodoxy: *Codelfa* and Ambiguity’ (n 4), which tracks this approach from the decision in *Shore v Wilson* (1842) 9 Cl & F 355; 8 ER 450.

¹¹ *Royal Botanic Gardens and Domain Trust v South Sydney City Council* (2002) 240 CLR 45, 62–3 [39] (*‘Royal Botanic Gardens and Domain Trust’*); *International Air Transport Association* (n 8) 160 [8]; *Western Export Services Inc v Jireh International Pty Ltd* (2011) 86 ALJR 1, 2 [2], 3 [6] (*‘Jireh’*); *Mount Bruce Mining Pty Ltd v Wright Prospecting Pty Ltd* (2015) 256 CLR 104, 116 [48]. See also *Rinehart v Welker* (2012) 95 NSWLR 221, 246 [115]–[116]; *Cordon Investments Pty Ltd v Lesdor Properties Pty Ltd* [2012] NSWCA 184, [52].

¹² *Jireh* (n 11). Although reasons for the resolution of a special leave to appeal application are not binding on lower courts.

Ltd v State Rail Authority of NSW [(1982) 149 CLR 337] by Mason J, with the concurrence of Stephen J and Wilson J, to be the ‘true rule’ as to the admission of evidence of surrounding circumstances. Until this Court embarks upon that exercise and disapproves or revises what was said in *Codelfa*, intermediate appellate courts are bound to follow that precedent. The same is true of primary judges, notwithstanding what may appear to have been said by intermediate appellate courts.

The position of *Codelfa*, as a binding authority, was made clear in the joint reasons of five Justices in *Royal Botanic Gardens and Domain Trust v South Sydney City Council* [(2002) 240 CLR 45] and it should not have been necessary to reiterate the point here.¹³

Notwithstanding the resolute statement in *Jireh* concerning the correctness of the ambiguity gateway, the controversy concerning the extent to which contextual surrounding circumstances are available, in the absence of ambiguity, to assist in the process of contractual interpretation remains alive. This is because intermediate appellate courts¹⁴ have read certain decisions of the High Court of Australia¹⁵ as implicitly overruling the ambiguity gateway. The principal reason for this reading of High Court authorities is that the High Court appears willing to consider non-notorious background facts to resolve contractual interpretation disputes absent a finding of ambiguity. For example, the High Court has used the contextual clue that a promisee under a long-term commodity supply agreement knew that the promisor had other customers when interpreting a ‘best endeavours’ clause.¹⁶ On this view there is no ‘ambiguity gateway’ rule where justices of the High Court implicitly say, ‘ambiguity gateway for thee but not for me’. This is because the ratio decidendi of any decision where the High Court fails to apply such a rule will also bind lower courts. However, not all legal commentators,¹⁷ nor all intermediate appellate courts share the view that the High Court has implicitly overruled the ambiguity gateway.¹⁸ The controversy in this area remains alive, with differently constituted intermediate

¹³ *Jireh* (n 11) 2–3 [4]–[5].

¹⁴ *Mainteck Services Pty Ltd v Stein Heurtey SA* (2014) 89 NSWLR 633, 653 [71]–[73] (‘*Mainteck*’); *Stratton Finance Pty Ltd v Webb* (2014) 314 ALR 166, 173–4 [36]–[41] (‘*Stratton Finance*’); *Newey v Westpac Banking Corporation* [2014] NSWCA 319, [89] (‘*Newey*’); *WIN Corporation Pty Ltd v Nine Network Australia Pty Ltd* (2016) 341 ALR 467, 478 [59]; *Cherry v Steele-Park* (2017) 96 NSWLR 548, 566 [76] (‘*Cherry*’); *Promoseven Pty Ltd v Markey* (2015) 104 ACSR 384, 408–9 [98]–[99].

¹⁵ See especially *Electricity Generation Corporation v Woodside Energy Ltd* (2014) 251 CLR 640, 656–7 [35] (‘*Woodside*’). See also the authorities collected in *Stratton Finance* (n 14) 173 [37] and *Lion Nathan Australia Pty Ltd v Coopers Brewery Ltd* (2006) 156 FCR 1, 10–13 [45]–[56]; 28 [122]; 48 [238]; and the discussion in Richard Calnan, *Principles of Contractual Interpretation* (Oxford University Press, 2013) 47.

¹⁶ *Woodside* (n 15) 656–7 [35].

¹⁷ See Prince, ‘Defending Orthodoxy: *Codelfa* and Ambiguity’ (n 4) 499; JD Heydon, ‘Comment on Lord Hoffmann’s “Interpretation of Contracts”’ in John Sackar and Thomas Prince (eds), *Heydon: Selected Speeches and Papers* (Federation Press, 2018) 710, 718.

¹⁸ For authorities that are less sanguine that the ambiguity gateway has been abolished, see, eg, *Technomin Australia Pty Ltd v Xstrata Nickel Australasia Operations Pty Ltd* (2014) 48 WAR 261, 271 [45]; 298–9 [212]–[217] (‘*Technomin Australia*’); *Watson v Scott* [2016] 2 Qd R 484, 495 [30]; *Apple and Pear Australia Ltd v Pink Lady America LLC* (2016) 343 ALR 112, 155 [137]–[138], 178–9 [231]–[232]; *Siemens Gamesa Renewable Energy Pty Ltd v Bulgana Wind Farm Pty Ltd* [2020] VSC 126, [99] (‘*Siemens Gamesa*’).

appellate courts taking different views on the issue.¹⁹ It is important to observe that the issue of the desirability of the ambiguity gateway ultimately will not be resolved by narrow arguments as to whether the High Court has implicitly overruled itself. Rather, the issue will need to be resolved by squarely addressing the potential justifications for the ambiguity gateway and, in turn, assessing to what extent the gateway detracts from the process of contractual interpretation.²⁰

Removing the ambiguity gateway would align Australian law closely with the more liberal approach adopted in England and Wales as enunciated by Lord Hoffmann in the ‘celebrated’²¹ decision in *Investors Compensation Scheme Ltd v West Bromwich Building Society*.²² That approach being that contractual language should be read in the first instance against its full set of background facts (absent ambiguity). Provided the facts are reasonably available to the parties and are relevant to establishing how a reasonable person would understand what the parties intended by the language used. Although commentators and judges have noted a recent emerging judicial trend in England and Wales away from the principles enunciated in *Investors Compensation Scheme* and towards a greater focus on contractual text,²³ that trend should not be overstated, as Lord Hoffmann’s principles are yet to be overruled.²⁴ A similar approach to that enunciated by Lord Hoffmann in *Investors Compensation Scheme* has been adopted in many other Commonwealth jurisdictions.²⁵ For example, an equivalent principle has been endorsed by the Supreme Court of Canada,²⁶ the Supreme Court of New Zealand,²⁷ the Court of

¹⁹ See *Aurizon Network* (n 5).

²⁰ See also James Edelman, ‘The Interpretation of Written Contracts’ in Charles Mitchell and Stephen Watterson (eds), *The World of Maritime and Commercial Law: Essays in Honour of Francis Rose* (Hart Publishing, 2020) 243.

²¹ See, eg, Gerard McMeel, ‘Foucault’s Pendulum: Text, Context and Good Faith in Contract Law’ (2017) 70(1) *Current Legal Problems* 365, 368. See also Stevens (n 3) 167, 174–8; Calnan (n 15) 45.

²² *Investors Compensation Scheme Ltd v West Bromwich Building Society* [1998] 1 WLR 896, 912–13 (‘*Investors Compensation Scheme*’).

²³ See McMeel (n 21); and Lord Jonathan Sumption, ‘A Question of Taste: The Supreme Court and the Interpretation of Contracts’ (2017) 17(2) *Oxford University Commonwealth Law Journal* 301, for example, noting decisions such as *Arnold v Britton* [2016] AC 1619; *Krys v KBC Partners* [2015] UKPC 46; *Marks & Spencer plc v BNP Paribas Securities Services Trust Co* [2016] AC 742. A useful summary of many of these authorities can be found in Ryan Catterwell ‘Striking a Balance in Contract Interpretation: The Primacy of the Text’ (2019) 23(1) *Edinburgh Law Review* 52.

²⁴ Sumption (n 23). For a discussion of the disagreement between Lord Sumption and Lord Hoffmann, see Ewan McKendrick, ‘Interpretation’ in William Day and Sarah Worthington (eds) *Challenging Private Law: Lord Sumption on the Supreme Court* (Hart Publishing, 2020) 3.

²⁵ On Australia becoming out of step with the rest of the common law world, see *Mainteck* (n 14) 655–6 [84].

²⁶ See, eg, *Sattva Capital Corp v Creston Moly Corp* [2014] 2 SCR 633, 656–8 [46]–[48]; *Uniprix inc v Gestion Gosselin et Bérubé inc* [2017] 2 SCR 59, 79–82 [35]–[41].

²⁷ See, eg, *Vector Gas Ltd v Bay of Plenty Energy Ltd* [2010] 2 NZLR 444, 457–9 [19]–[22]; *Firm PI 1 Ltd v Zurich Australian Insurance Ltd* [2015] 1 NZLR 432, 453–5 [60]–[63]. Although note subsequent pronouncements of the same Court, which place somewhat more emphasis on the plain meaning of the communicative act creating the contract: *Lakes International Golf Management Ltd v Vincent* [2017] 1 NZLR 935, 944–6 [23]–[30].

Appeal of the Republic of Singapore²⁸ and the Hong Kong Court of Final Appeal.²⁹ Of course, being an outlier does not, in and of itself, demonstrate that the law in Australia has taken a wrong turn. After all, the group can all too often get something wrong and a minority of one can be right. Indeed, whether Australia is in a minority depends upon the sample selected. If the eye is cast beyond Commonwealth jurisdictions, for example, it will be observed that most states in the United States of America ('US') have maintained a rule similar to the ambiguity gateway.³⁰ The salient point to take from the discussion above provides the relevant context for a point of law that is ripe soon to be considered by the High Court of Australia:³¹ whether recourse to the surrounding circumstances accessible to the parties at the time of entry into the contract should be permissible absent ambiguity.

III Justifications – Do You Prefer Accuracy or Utility?

A *Sentence Meaning, Speaker Meaning and Objectivity in Interpretation*

The central reason why ambiguity should not be required before having recourse to relevant and probative contractual context is that it detracts from the interpretive process. To understand why this is so, some space needs to be dedicated to painting a brief picture of how language and communication operate. During the 20th century, developments in the philosophy of language and mind resulted in an understanding of human communication that depends on external 'rules' to divine the intentions of an author of an utterance. This understanding of human communication draws a

²⁸ See, eg, *Zurich Insurance (Singapore) Pte Ltd v B-Gold Interior Design & Construction Pte Ltd* [2008] 3 SLR 1029, 1087–92 [114]–[124]; *Sembcorp Marine Ltd v PPL Holdings Pte Ltd* [2013] 4 SLR 193, 224–6 [72]–[74] ('*Sembcorp Marine*'). The decision in *Sembcorp Marine* is discussed in more detail in the text accompanying n 119 below.

²⁹ See, eg, *Fully Profit (Asia) Ltd v Secretary for Justice* (2013) 16 HKCFAR 351, 361 [15].

³⁰ Those US states favouring the ambiguity gateway being often termed 'Willistonian' after Samuel Williston and those favouring a wider approach to context often termed 'Corbinian' after Arthur Corbin: see Brian H Bix, *Contract Law: Rules, Theory, and Context* (Cambridge University Press, 2012) 61. See also Schwartz and Scott, who note '[n]ine states, joined by the Uniform Commercial Code for sales cases (UCC) and the Restatement (Second) of Contracts, have adopted a contextualist [ie, one with no ambiguity gateway] ... interpretative regime': Alan Schwartz and Robert E Scott, *Contract Interpretation Redux* (2010) 119(5) *Yale Law Journal* 926, 928. New York is the most significant commercial jurisdiction that preserves the ambiguity gateway, whereas California is the most significant commercial jurisdiction to abolish the gateway.

³¹ Being a standard reason why special leave to appeal to the High Court of Australia is granted: see, eg, *Judiciary Act 1903* (Cth), s 35A; Justice Michael Kirby, 'Maximising Special Leave Performance in the High Court of Australia' (2007) 30(3) *University of New South Wales Law Journal* 731, 743.

formal distinction between the ‘speaker’³² and the ‘sentence or conventional’³³ meaning of an utterance.³⁴

Consider the following infamous newspaper extract: ‘Yoko Ono will talk about her husband John Lennon who was killed in an interview with Barbara Walters’.³⁵ The communication is capable of bearing at least two meanings: (i) a narrow ‘sentence meaning’ whereby the text informs the reader that Barbara killed John (or at the very least John was killed whilst Barbara was interviewing him); or (ii) a more contextual ‘speaker’ meaning whereby the journalist is ‘most likely’ intending to inform the reader that Yoko will be discussing the murder of John in an interview with Barbara. Thus, what is immediately evident is that the recognition of notorious background context (for example, that Barbara is a high profile broadcast journalist) and the recipients’ powers of rationality in the interpretive process give the text a meaning different from its sentence meaning. This is because human beings do not communicate merely by virtue of the ‘sentence meaning’ of an utterance alone: that is, by a process solely of decoding a message in light of specific narrow customary rules.

Rather, communication involves a process that is inferential; it involves inductive and not deductive reasoning.³⁶ The central point of interpretation is to infer

³² While the speaker meaning and sentence meaning of words will often coincide, they can come apart. For example, I am at a café and order a Vienna coffee. I am, however, unaware that a Vienna coffee contains cream. I mistakenly believe that I am ordering a coffee without cream. In this example, I have misused a word, I intended Vienna coffee to mean a coffee without cream and I also intended to be taken by the barista to have intended so (although my usage of the word Vienna coffee will not, without more, be understood as such by the barista as my usage was unconventional). In this example, my mistaken reference to Vienna coffee can be termed the ‘speaker meaning’ to be ascribed to the word. On speaker meaning, see David Goddard, ‘The Myth of Subjectivity’ (1987) 7(3) *Legal Studies* 263, 265–6; Richard Ekins and Jeffery Goldsworthy, ‘The Reality and Indispensability of Legislative Intentions’ (2014) 36(1) *Sydney Law Review* 39, 47; Robin Kar and Margaret Radin, ‘Pseudo-Contract and Shared Meaning Analysis’ (2019) 132(4) *Harvard Law Review* 1135, 1145–6.

³³ The sentence meaning being the meaning ascribed to an utterance by use of a conventional standard. Searle notes that it is ‘the creation of conventional devices for performing acts of speaker meaning, which gives us something approaching sentence meaning, where sentence meaning is the standing possibility of speaker meaning. Sentence meaning is conventionalized.’: John R Searle, ‘What is Language? Some Preliminary Remarks’ (2009) 11(1) *Ethics & Politics* 173, 193 (‘What is Language?’).

³⁴ In this article, I use the term ‘utterance’ to mean ‘communicative act’, for example, writing also constitutes an ‘utterance’.

³⁵ The example is taken from Steven Pinker, *The Language Instinct: How the Mind Creates Language* (Penguin Science, 1994) 102.

³⁶ Ludwig Wittgenstein, *Philosophical Investigations* (Wiley-Blackwell, 4th ed, 2009) 86–111 [198]–[315]; John L Austin, *How To Do Things with Words* (Clarendon Press, 1962); John L Austin, ‘Performative Utterances’ in John L Austin, *Philosophical Papers* (Oxford University Press, 2nd ed, 1970) 233; Saul Kripke, ‘Speaker’s Reference and Semantic Reference’ (1977) 2(1) *Midwest Studies in Philosophy* 255; John R Searle, *Expression and Meaning: Studies in the Theory of Speech Acts* (Cambridge University Press, 1979); Saul Kripke, *Wittgenstein on Rules and Private Language* (Blackwell Publishing, 1982); Paul Grice, *Studies in the Way of Words* (Harvard University Press, 1989); Michael Dummett, *The Seas of Language* (Clarendon University Press, 1993) 97–105; John R Searle, *Mind, Language and Society: Philosophy in the Real World* (Basic Books, 1999) 139–45, especially at 144–5:

Grice saw correctly that when we communicate to people, we succeed in producing understanding in them by getting them to recognize our intention to produce that understanding. Communication is peculiar among human actions in that we succeed in producing an intended effect on the hearer by getting the hearer to recognize the intention to produce that very effect.

the author's most probable intention from the communicative act. As such, intentionality provides a guide in this process.³⁷ A recipient of an utterance will consider what it means by inductively balancing the competing rules and principles through which intentionality has been funnelled; namely, the public meaning of the specific words the author has deployed and a range of contextual factors — for example, the assumed existence of shared background information, a recipient's general powers of reasoning and rationality, and that parties to a conversation intend to communicate meaningfully. Kripke has made this point in the following terms:

The notion of what words can mean, in the language, is semantical: it is given by the conventions of our language. What they mean, on a given occasion, is determined, on a given occasion, by these conventions, *together with the intentions of the speaker and various contextual features*. Finally what the speaker meant, on a given occasion, in saying certain words, derives from various *further special intentions of the speaker, together with various general principles, applicable to all human languages regardless of their special conventions*.³⁸

Consider an example where Dixon asks Frankfurter to go to the 'Eagle & Child Public House tonight for a meal at 6pm'. Frankfurter responds: 'I have a train to catch'. Frankfurter's response is generally understood to mean that he is rejecting Dixon's proposal, but this cannot be explained by virtue of the narrow sentence meaning of the text or utterance alone. The reasoning deployed to take Frankfurter's utterance as a rejection of Dixon's proposal appears to be that:

- (i) it is a rule of interpretation that, unless there is evidence to the contrary, the recipient (Dixon) assumes that the author (Frankfurter) is attempting to communicate meaningfully and cooperate in the conversation (that is, a starting rule that Frankfurter is not speaking nonsense);³⁹
- (ii) from Dixon's perspective it appears that Frankfurter must have meant something more than the literal meaning of what he said as the literal meaning of the words neither expressly reject nor accept the proposal to go to the pub;
- (iii) Dixon (and the average person for that matter) understands certain notorious background 'contextual' information (such as that one cannot be in two places at once, a train will run on limited schedules and tickets can be non-refundable etc);
- (iv) given the content of (iii), then the rational person⁴⁰ in Dixon's position will realise that it is unlikely that Frankfurter can both: (a) attend the pub; and (b) catch his train;

... I am trying to tell someone that it is raining, I succeed in telling them as soon as they recognize that I am trying to tell them.

³⁷ I will put to one side the issue of how corporate bodies have intentions. Others have grappled aptly with this issue. On the collective intentions of non-natural persons (eg corporations and legislatures) and collective intentionality in general, see Ekins and Goldsworthy (n 32) 47; Ryan Catterwell, *A Unified Approach to Contract Interpretation* (Hart Publishing, 2020) 92 [4-20].

³⁸ Kripke, 'Speaker's Reference and Semantic Reference' (n 36) 263 (emphasis added).

³⁹ See also Kar and Radin (n 32) 1147-50.

⁴⁰ Thus, Lord Hoffmann was correct to observe in *Investors Compensation Scheme* (n 22) 913: 'Many people, including politicians, celebrities and Mrs Malaprop, mangle meanings and syntax but

(v) given that to accept a proposal one must be able to perform his side of whatever the proposal is, as a matter of basic inductive reasoning⁴¹ it appears to be most probable that Frankfurter is rejecting (politely, by saying so indirectly) Dixon's proposal as he has limited capacity to attend the pub and his communication, which was made in direct response to a proposal, likely has a purpose.⁴²

Contextual and purposive reasoning is standard in both Australian and English contractual interpretation jurisprudence.⁴³

The critical reader at this point may respond along these lines: 'it is fine that you have identified a particular development in the philosophy of language and mind between sentence meaning and speaker meaning, but how does this impact on what lawyers do'? This is a fair critique; it can often be doubted whether specialised philosophical arguments reflect the messy way in which the judge-made law develops in both a diffuse and incremental manner.⁴⁴ My response, however, is that the distinction drawn above between the sentence and speaker meaning of an utterance is crucial for understanding Lord Hoffmann's speech in *Investors Compensation Scheme*. As his Lordship said:

The meaning which a document (or any other utterance) would convey to a reasonable man is not the same thing as the meaning of its words. The meaning of words [that is, the sentence meaning in the sense used above] is a matter of dictionaries and grammars; the meaning of the document is what the parties using those words against the relevant background would reasonably have been understood to mean. The background may not merely enable the reasonable man to choose between the possible meanings of words which are ambiguous but even (as occasionally happens in ordinary life) to conclude that the parties must, for whatever reason, have used the wrong words or syntax.⁴⁵

The point is that the general law does not fix upon the sentence meaning of a contract (that is, the narrow meaning of the words read in artificial isolation). Rather, it asks

nevertheless communicate tolerably clearly what they are using the words to mean. If anyone is doing violence to natural meanings, it is they rather than their listener.'

⁴¹ Dummett (n 36) 104 (emphasis in original): 'Any adequate philosophical account of language must describe it as a rational activity on the part of creatures to whom can be ascribed *intention* and *purpose*.'

⁴² This idea has been expanded on by relevance theory, which argues that the meaning of express linguistic expressions is generally underdetermined such that there is a significant gap between the intentions of the speaker and narrow literal meaning of an utterance. See, eg, Deirdre Wilson, 'Relevance Theory' in Yan Huang (ed) *The Oxford Handbook of Pragmatics* (Oxford University Press, 2017) 79, 85–9; Robyn Carston, *Thoughts and Utterances: The Pragmatics of Explicit Communication* (Blackwell Publishing, 2002) 83.

⁴³ There are plenty of examples, but four illustrative examples are *Royal Botanic Gardens and Domain Trust* (n 11) 62 [36]; *Woodside* (n 15) 660–2 [44]–[50]; *Ecosse Property Holdings Pty Ltd v Gee Dee Nominees Pty Ltd* (2017) 261 CLR 544, 551 [16], 555 [27]; *Thorney Park Golf v Myers Catering Ltd* [2015] EWCA Civ 19, [26] ('*Thorney Park Golf*').

⁴⁴ See, eg, Nicholas J McBride, *The Humanity of Private Law: Part I: Explanation* (Hart Publishing, 2018) 27–28, 261, 264–5.

⁴⁵ *Investors Compensation Scheme* (n 22) 913. See also *Kirin-Amgen Inc v Hoechst Marion Roussel Ltd* [2005] 1 All ER 667, 689 [64] (Lord Hoffmann). On Lord Hoffmann's use of philosophy of language, see Paul S Davies, 'The Meaning of Commercial Contracts' in Paul S Davies and Justine Pila (eds) *The Jurisprudence of Lord Hoffmann* (Hart Publishing, 2017) 215.

the following question: what the reasonable recipient of a legally significant communicative act would infer to be the author's most probable intention from that communicative act. As such, intentionality provides a guide in this process. The court arrives at the 'best answer' to this question by inductively balancing competing principles through which intentionality has been funnelled; namely, the public meaning of the specific words the author has deployed and a range of permissible contextual factors.⁴⁶ Context can, at its broadest, include: notorious background facts; prior negotiations and preparatory works; the purpose and internal logic of a written instrument; common industry and institutional practice; the parties' powers of rationality; and even shared normative understandings.⁴⁷ Although, there is a legitimate debate to be had concerning when context should yield to considerations of legal certainty and efficiency.⁴⁸

On this approach, the general law still cares about speaker intentionality, but it does so only objectively. The fact that the principles of contractual interpretation care about intentionality is evident in the following basal principles:

- (i) courts care about more than the literal meaning of the words of a contract read in artificial isolation (that is, sentence meaning in the sense used in this article does not govern the modern approach to contractual interpretation);
- (ii) courts will consider the purpose and object of the transaction (that is, the words of a contract do not have an abstract purpose, rather only parties to the contract have such a purpose);
- (iii) contracts are to be interpreted in light of the local context within the document (that is, the words of a contract do not have a context in and of themselves, rather the local context within the contract is relevant as the parties are taken to have intended to create, where possible, an internally logical and coherent agreement);
- (iv) contracts are to be interpreted in light of commercial common sense (that is, again there is nothing 'commercial' about the sentence meaning of words, rather a commercial interpretation is given to a contract where possible because viability is more likely to reflect what the author(s) intended); and
- (v) a large number of maxims of contractual interpretation (and interpretation more generally) can be understood only on the basis of intentionality.

Consider the following three maxims of contractual interpretation. First, the maxim *ejusdem generis* (of the same kind) turns on an author's most likely intention. If I say at the end of a lecture that my next lecture will be on 'the sun, the moon, the planets in our solar system, and other tremendous bodies' you infer that I do not intend 'other tremendous bodies' to include, for example, a detailed analysis of

⁴⁶ For good lists of contextual factors and related rules of interpretation see, eg, Stephen Smith, *Contract Theory* (Oxford University Press, 2004) 274; Sir George Leggatt, 'Making Sense of Contracts: the Rational Choice Theory' (2015) 131 (July) *Law Quarterly Review* 454, 468–70; Catterwell (n 37) 123–32 [5–31]–[5–59].

⁴⁷ Perhaps best reflected in a principle such as the *contra proferentem* rule.

⁴⁸ In this connection, see Schwartz and Scott arguing in favour of a more *Codelfa*-style approach to contractual interpretation that limits the use of extrinsic material in the United States of America ('US') on the basis of utility *albeit at the expense of interpretive accuracy*: Schwartz and Scott (n 30) 930.

catwalk models.⁴⁹ This is because you infer that I most likely intend to refer to *celestial* bodies given the context in which I chose to use the words and that my lecture is more likely than not to have a coherent theme. Second, the maxim *nosscitur a sociis* (it is known by its associates) is explained by an author's most likely intention. If I say you should use a 'case or a steel canister to carry explosives', you will infer that my general reference to 'case' needs to be read in light of my specific reference to 'steel canister'. You reach this conclusion by inferring that I do not mean 'case' to, for example, include a carrying case made of cardboard but rather a 'case' with similar characteristics to the steel canister given the purpose of my communication is to convey safety advice.⁵⁰ Third, the maxim *expressio unius est exclusio alterius* (the expression of one is the exclusion of others) is based on an author's presumed intentions. That is, an express inclusion of one or more things of a particular type in a communicative act necessarily implies an intention to exclude other things of that type. For example, if I agree to wash your windows this weekend and provide a detailed list of the services that I will render, then it can be inferred that I intended the services *not* on the list (say, for example, washing the windows of your car) to be excluded from our contract. This is for a simple reason based on presumed intentionality: why else would I have gone to such effort to list the particular services in the first place.⁵¹

The next fair critique of the argument that I have presented thus far would be: if the principles of contractual interpretation took intentionality seriously why is the process of interpretation objective and not subjective? That is, legal interpretation differs from day-to-day interpretation in that the former is objective and the latter subjective. My response would be as follows: where a party to a contract uses conventional standards (that is, the sentence meaning of an utterance taken with the standard contextual 'rules' or 'clues' to discern speaker meaning) to effect a promise with a counterparty, the promisor cannot now renege from the objective meaning attributed to her utterance by application of those conventions without damaging the shared legal institutions that those standards create. In short, there exist public conventions as to how the promisor can express her intentions.⁵² The use of such conventions creates expectations in others.⁵³ If one is to take the benefits of such conventions in order to enhance one's own autonomy to create contracts, then one

⁴⁹ See, eg, *Powell v Kempton Park Racecourse Co Ltd* [1899] AC 143.

⁵⁰ See, eg, *Foster v Diphwys Casson Slate Co* (1887) 18 QBD 428.

⁵¹ See, eg, *North Stafford Steel, Iron and Coal Co (Burslem) Ltd v Ward* (1868) LR 3 Exch 172.

⁵² See Goddard (n 32) 268–71.

⁵³ I have defended this position in more detail elsewhere: NA Tiverios, 'A Uniform Hermeneutic Thesis: Objectivity and the High Court of Australia's Approach to Interpretation across the Private Law' (2021) 40(2) *University of Queensland Law Journal* 181. See also Charles Fried, *Contract as Promise: A Theory of Contractual Obligation* (Harvard University Press, 1981) 12–13, 87–8; Searle, 'What is Language?' (n 33) 199–201, especially 199 (emphasis added):

We have to assume that [homo sapiens or some equivalent hypotheticalal species] are capable of evolving procedures for representing [internal] states of affairs; where the representations have speaker meaning They can represent states of affairs that they believe exist, states of affairs they desire to exist, states of affairs they intend to bring about, etc. . . . These procedures, or at least some of them, become conventionalized. What does that mean exactly? It means that given collective intentionality, if anyone intentionally engages in one of these procedures, then other members of the group have a right to expect that the procedures are being followed correctly. This, I take it, is the essential thing about conventions. *Conventions are arbitrary, but once they are settled they give the participants a right to expectations.*

must also take the burden that, as a matter of parity, others are entitled to those same benefits. The conventions of language and communication will break if employed disingenuously or incorrectly such that it is wholly justifiable to hold a promisor to the objective meaning attributed to her utterance.⁵⁴ This argument, however, is not intended to provide an indefeasible argument for a promisor always to do what was promised. There are rules of law and equity that recognise that there are other events of greater normative pull that can, in limited circumstances, justify releasing a promisor from the objective meaning attributed to her promise.⁵⁵

So understood, the background context to which the ambiguity gateway prevents recourse is an additional clue that *may* prove useful in inferring the objective intentions of the author(s) of a contract. As Edelman J observed in *Rinehart v Hancock Prospecting Pty Ltd*: ‘Every clause in a contract, no less arbitration clauses, must be construed in context. No meaningful words, whether in a contract, a statute, a will, a trust, or a conversation, are ever acontextual.’⁵⁶ Grice expressed the same point as follows:

[I]n cases where there is doubt, say, about which of two or more things an utterer intends to convey, we tend to refer to the context (linguistic or otherwise) of the utterance and ask which of the alternatives would be relevant to other things he is saying or doing, or which intention in a particular situation would fit in with some purpose he obviously has (e.g., a man who calls for a ‘pump’ at a fire would not want a bicycle pump). Nonlinguistic parallels are obvious: context is a criterion in settling the question of why a man who has just put a cigarette in his mouth has put his hand in his pocket; relevance to an obvious end is a criterion in settling why a man is running away from a bull.⁵⁷

If this is the case, depriving the court of otherwise probative and relevant surrounding circumstances as a means to understand language used in a contract detracts from this interpretive process. The ambiguity gateway disables the court from having available all the relevant context to reach the ‘best possible’ interpretation of a legally significant communicative act.⁵⁸ The best possible interpretation being that inferred by the reasonable recipient as the author’s most probable intention from that communicative act. Indeed, on my argument in this article, there is much force in arguments that courts should, in general, be more

⁵⁴ This argument mirrors Kant’s famous example that in a society where the truth of an expression can no longer be taken at face value, the conventional standard of promising would be swiftly abolished: Immanuel Kant, *Groundwork of the Metaphysics of Morals* (Cambridge University Press, 2017 rev ed) 17–19 [4:402]–[4:403]. Note that other normative theories come to a conclusion not dissimilar to that I adopt in this article: see Joseph Raz, ‘Review: Promises, Morals, and Law’ (1982) 95(4) *Harvard Law Review* 916, 936–8 (justifying objectivity based on utilitarianism on the grounds that it protects the institution of promising from harm); John Finnis, *Natural Law and Natural Rights* (Oxford University Press, 2nd ed, 2011) 303 (justifying objectivity based on the stability and cooperation required to build the ‘common good’ from the perspective of natural law). For a view of natural law like Finnis, see McBride (n 44) 165: ‘Contract law would fail in its mission to facilitate the orderly workings of the marketplace were it *not* to give effect to the objective principle’.

⁵⁵ Obvious examples include the doctrine of restraint of trade, relief against penalties, relief against forfeiture, and vitiating factors.

⁵⁶ *Rinehart v Hancock Prospecting Pty Ltd* (2019) 267 CLR 514, 548 [83]. See also below n 91.

⁵⁷ Paul Grice, ‘Meaning’ (1957) 66(3) *Philosophical Review* 377, 387.

⁵⁸ See also Carter (n 3) 118; McLauchlan, ‘Contractual Interpretation: What Is It About?’ (n 3) 31–5, where McLauchlan notes further arguments based on: (i) coherence in the law; (ii) transparency; and (iii) coherence with international law.

sanguine as to the use of evidence of prior negotiations⁵⁹ as providing further context to discern the meaning that the reasonable recipient of the communication would attribute to the contract.⁶⁰ The State should only enforce contractual rights, duties, powers, liabilities, privileges and immunities between A and B that those parties have objectively intended to have assented to. Accuracy in contractual interpretation helps to facilitate this goal.

B Justifications(?) for the Ambiguity Gateway

Given the above conclusion, what could be said in defence of the ambiguity gateway? The most common explanation is that it is potentially justifiable on the basis of classic rule-based utilitarianism, where both (i) the interests of parties in enforcing and drafting contracts as *an entire class*; and (ii) the State's role (and limited resources) as an umpire in the enforcement of voluntary exchanges, are protected.⁶¹ That is, the principle exists to make the resolution of contractual disputes by the State more efficient⁶² *in globo*.⁶³ Put simply, the basic arithmetic is that where a contract is unambiguous, the marginal gains in accuracy in the interpretive process by considering the full set of relevant surrounding circumstances in the first instance are outweighed by the burdens associated with having such evidence relevant by default. This is particularly the case if, in any event: (i) the evidence of the surrounding circumstances is unlikely to change the meaning of the contract where the contractual language is unambiguous; or (ii) enough contextual information has been included within the contract itself or supplied by the notorious background facts in order to reach the correct interpretation.⁶⁴

As an alternative to this classical approach to utilitarianism as a moral philosophy, a more modern form of utilitarian reasoning has been adopted in US scholarship to justify the ambiguity gateway. This approach derives utility not by

⁵⁹ For limits on the use of prior negotiations in contractual interpretation, see *Byrnes v Kendle* (2011) 243 CLR 253, 285 [99] (Heydon and Crennan JJ); *Investors Compensation Scheme* (n 22) 912–13.

⁶⁰ See the forceful argument made in David McLauchlan, 'The Continuing Confusion and Uncertainty over the Relevance of Actual Mutual Intention in Contract Interpretation' (2021) 37(1–2) *Journal of Contract Law* 25. See also *Cherry* (n 14) 569 [91]–[92].

⁶¹ Smith (n 46) 275; Catherine Mitchell, *Interpretation of Contracts* (Routledge-Cavendish, 2nd ed, 2007) 79. See also the discussion in Leggatt (n 46) 465.

⁶² By this I mean Kaldor-Hicks efficient (the rule produces more benefits than costs *in globo*), rather than Pareto efficient (the rule makes *everyone* better off). Law and economics scholarship typically uses Kaldor-Hicks efficiency: Smith (n 46) 110–11.

⁶³ See, eg, Spigelman (n 4) 334; Prince, 'Defending Orthodoxy: *Codelfa* and Ambiguity' (n 4) 503–9. See also Jonathan Morgan, *Contract Law Minimalism* (Cambridge University Press, 2013) 233: 'Rules in commercial matters should be as clear as possible to enable decisions to be made swiftly and confidently'. Although, in making this point Morgan is cognisant of the insights from philosophy of language and the mind noted in this article.

⁶⁴ Sumption (n 23). Lord Sumption criticised the principles enunciated in *Investors Compensation Scheme* (n 22) 912–13, on the basis that rather than being used to *interpret* the language, the principles are often deployed by judges to consider the *reasonableness* of the contract: Sumption (n 23). See also the authorities collected in Kim Lewison, *The Interpretation of Contracts* (Sweet and Maxwell, 6th ed, 2015) 14 [1.04]; Prince, 'Defending Orthodoxy: *Codelfa* and Ambiguity' (n 4) 501. See further Schwartz and Scott, who note that 'parties will include contextual bits until the marginal gain—the increased expected contractual payoff [in interpretative accuracy] from further bits equals the marginal cost of writing them': Schwartz and Scott (n 30) 954.

virtue of the traditional Benthamic weighing up of costs and benefits, but by assessing utility from consistent individual choices.⁶⁵ Such scholars argue that sophisticated commercial parties prefer the ambiguity gateway⁶⁶ and that the default rules of contractual interpretation should reflect the choices such individuals would typically make for themselves (albeit leaving specific parties free to choose whether or not to contract around the default rule).⁶⁷ More will be said about these arguments below.⁶⁸ I will initially address the classical rule-based utilitarian reasoning as that is the type of analysis typically found in Anglo-Australian contract law literature and jurisprudence.

Arguing in favour of a more textual (that is, *Codelfa*-style) approach to contractual interpretation in the US, Schwartz and Scott have said:

although accurate judicial interpretations are desirable, accurate interpretations are costly for parties and courts to obtain. If contract writing were free, parties could minimize interpretive error by exhaustively detailing their intentions. And if adjudication were costless, courts could minimize interpretive error by hearing all relevant and material evidence. Contract writing and litigation are costly, however. Since no interpretive theory can justify devoting infinite resources to achieving interpretive accuracy, any socially desirable interpretive rule would trade off accuracy against contract-writing and adjudication costs. Such a rule, we argue, tells courts in some cases to exclude relevant evidence.⁶⁹

Such concerns are perfectly understandable given that any commercial lawyer is well aware that examples of an overabundance of voluminous trial bundles in the course of litigation are legion.⁷⁰ In one illustrative case, Simon Bryan QC (sitting as a Deputy High Court Judge) noted the regrettable inclusion in the trial bundles of ‘no less than eight lever arch files full of what were described, somewhat unpromisingly, as “Draft Contractual Documents”’.⁷¹ This is a scene repeated often throughout the common law world. Indeed, I suspect there will be some practitioners reading this who will consider eight lever arch files of ‘context’ to be an example of virtuous restraint.

There is much force in the idea that legal practitioners and judges should be slow to decry the problems pertaining to access to justice and the rising costs of litigation one day and then incrementally, and for what may prove to be a limited benefit in the name of an elusive search for perfect individualised justice, proceed to make the law a little more complex the next day. In every contractual dispute, having

⁶⁵ Ken Binmore, ‘Interpersonal Comparison of Utility’ in Don Ross and Harold Kincaid (eds), *The Oxford Handbook of Philosophy of Economics* (Oxford University Press, 2009) 540, 542–9; Doug Bernheim, ‘Behavioral Welfare Economics’ (2009) 7(2–3) *Journal of the European Economic Association* 267; Hanoch Dagan and Roy Kreitner, ‘Economic Analysis in Law’ (2021) 38(2) *Yale Journal on Regulation* 566, who note at 572 that ‘[s]tandard works in normative law and economics study the incentive effects of legal rules and doctrines and examine how they fare vis-à-vis the normative criterion of maximizing aggregate preference satisfaction’.

⁶⁶ Schwartz and Scott (n 30) 931.

⁶⁷ Bernheim (n 65) 291–2.

⁶⁸ See below text accompanying nn 96–116.

⁶⁹ Schwartz and Scott (n 30) 930.

⁷⁰ See also the similar concerns noted in Heydon (n 17) 720–1.

⁷¹ *BP Gas Marketing Ltd v La Societe Sonatrach* [2016] EWHC 2461 (Comm) [103].

recourse to the full set of surrounding circumstances in the first instance during the interpretive process would mean extra work and time devoted to providing legal advice, preparing a case for trial, advocacy, and writing a judgment. Moreover, I am yet to note the disbursements for producing bundles of trial documents. As cautioned in *Aon Risk Services Australia Ltd v Australian National University*:

While in general it is now seen as desirable that most types of litigation be dealt with expeditiously, it is commonly seen as especially desirable for commercial litigation. Its claims to expedition may be less than those of proceedings involving, for example, extraordinary prejudice to children; or the abduction of children; or a risk that a party will lose livelihood, business or home, or otherwise suffer irreparable loss or extraordinary hardship, unless there is a speedy trial. But commercial litigation does have significant claims to expedition. Those claims rest on the idea that a failure to resolve commercial disputes speedily is injurious to commerce, and hence injurious to the public interest. ...

The courts are thus an important aspect of the institutional framework of commerce. The efficiency or inefficiency of the courts has a bearing on the health or sickness of commerce.⁷²

What, then, should we make of arguments pertaining to legal efficiency (and we can include the closely related arguments regarding legal certainty)⁷³? Neither economic efficiency nor legal certainty alone is persuasive as a justification for a particular legal rule. This is because such arguments prove too much: arguments with a narrow focus on efficiency or certainty do not attempt to justify their claim solely to control the law.⁷⁴ They ‘beg the question’ in the proper sense of that phrase (that is, such analysis assumes the correctness of its underlying arguments without proving it). The position I take here is that a legal rule that is morally indefensible cannot be saved by recourse to arguments centring exclusively on either efficiency or certainty.⁷⁵ Let us look at two historical examples:

Example 1: A owes B a debt of \$5,000 payable on date X. Both A and B live in Darwin. A fails to pay B by date X. Indeed, A never pays B. Instead, A swears under oath that he paid B the \$5,000 and, in turn, 11 ‘witnesses’ situated in Sydney swear on oath as to A’s character. Imagine if the common law considers A’s oath as

⁷² *Aon Risk Services Australia Ltd v Australian National University* (2009) 239 CLR 175, 223 [137], citing *Collins v Mead* (Supreme Court of New South Wales, Rogers J, 7 March 1990) 220.

⁷³ See, eg, *Siemens Gamesa* (n 18) [99]. Brevity is the main reason why I have dealt with certainty and efficiency together. While some may quibble with me raising them together, I nonetheless appreciate that these are distinct (but related) concepts. As von Hayek notes in FA Hayek, *The Constitution of Liberty* (Routledge & Kegan Paul, 1960) 208:

The importance which the certainty of the law has for the smooth and efficient running of a free society can hardly be exaggerated. There is probably no single factor which has contributed more to the prosperity of the West than the relative certainty of the law which has prevailed here. This is not altered by the fact that complete certainty of the law is an ideal which we must try to approach but which we can never perfectly attain ...

⁷⁴ ‘[T]he vast majority of law and economics scholarship assumes without hesitation that the goal of the law should be efficiency’: Jon Hanson, Kathleen Hanson and Melissa Hart, ‘Law and Economics’ in Dennis Patterson (ed), *A Companion to Philosophy of Law and Legal Theory* (Wiley-Blackwell, 2nd ed, 2010) 300; see also at 322–4. Further, the theory of efficient breach has drawn some criticism from the High Court of Australia, see *Tabcorp Holdings Ltd v Bowen Investments Pty Ltd* (2009) 236 CLR 272, 285–90 [13]–[20].

⁷⁵ Stevens (n 3) 170.

supported by his ‘character witnesses’ as being good discharge of the debt. Readers familiar with English legal history will know that we do not have to imagine. Indeed, debt cases worked this way for some time. My example is no more than a modern take on the common law practice of wager of law or compurgation used by defendants in a simple debt case. Naturally, this approach is both a certain and efficient way of resolving a debt case. However, I doubt many readers would consider the oaths of 11 ‘witnesses’⁷⁶ with no local nexus to, or knowledge regarding, the dispute attesting to A’s credibility that he has paid B a debt is a sound way of resolving the substantive merits of a debt claim.⁷⁷

Example 2: B executes a deed saying that he will pay A the sum of \$5,000 on date X. B pays A the sum of \$5,000 on date X. A nonetheless remains in custody of the ‘physical’ deed. Imagine if the common law allowed A to enforce the deed against B and required B to pay to A a further \$5,000. Again, readers familiar with English legal history will know that we do not have to imagine. This is because common law procedure enabled an obligee holding a simple bond to enforce the bond multiple times.⁷⁸ Curiously, the justification for this rule is that the common law procedure favoured the universal benefits of simplicity, efficiency and certainty in making the mere production of the bond to the common law court constitute non-traversable proof of an obligation to pay a debt as stipulated in the bond. Serjeants Staunford and Bromley captured the utilitarian justification for double recovery:

it is nevertheless better to suffer mischief to one man than inconvenience to many, which would subvert the law. For if matter in writing could be so easily defeated and avoided by such a surmise, by naked breath, a matter in writing would be of no greater authority than a matter of fact.⁷⁹

Making the production of a bond in court non-traversable proof of a debt evidenced in that bond is a very certain and efficient way of resolving a debt case. However,

⁷⁶ Often paid witnesses based in London.

⁷⁷ See Sir John Baker, *An Introduction to English Legal History* (Oxford University Press, 5th ed, 2019) 81; Theodore FT Plucknett, *A Concise History of the Common Law* (Lawbook Exchange, 5th ed, 2010) 115–6. The process was fully abolished in 1833 by statute: see *Civil Procedure Act 1833*, 3 & 4 Will. IV, c. 42, s 13. On the history of bonds, see further, Nicholas A Tiverios, *Contractual Penalties in Australia and the United Kingdom: History, Theory and Practice* (Federation Press, 2019) 10–39.

⁷⁸ Indeed, in *Donne v Cornwall* (1486) YB Pass 1 Henry VII, Fo 14v, Pl 2 (CP), extracted in Sir John Baker, *Baker and Milsom Sources of English Legal History: Private Law to 1750* (Oxford University Press, 2nd ed, 2010) 283, A successfully sued in the Common Pleas and then on appeal to the non-statutory Exchequer Chamber on a simple bond that he stole back from B’s wife, after B had already paid to A the sum of £10 owing under the bond. Relief became available in equity during the reign of Edward IV (1442–83): Alfred WB Simpson, ‘The Penal Bond with Conditional Defeasance’ (1966) 82 (July) *Law Quarterly Review* 392, 416–18; DEC Yale (ed), *Lord Nottingham’s Chancery Cases (Vol 2)* (Bernard Quaritch, 1957–61) 9; DEC Yale (ed) *Lord Nottingham’s ‘Manual of Chancery Practice’ and ‘Prolegomena of Chancery and Equity’* (Cambridge University Press, 1965) 213; WT Barbour, *The History of Contract in Early English Equity* (Clarendon Press, 1914) 85–9; Theodore FT Plucknett and John L Barton (eds), *St Germain’s Doctor and Student* (Selden Society, 1974) 77–8. The label ‘utilitarian’ could be used to justify this approach, notwithstanding that term was not in use in the 15th century. Utilitarianism would only become an identifiable moral and political philosophy after the 1780s, with the publication (1789) of Jeremy Bentham’s *An Introduction to the Principles of Morals and Legislation* (although the nomenclature of ‘utility’ was borrowed by Bentham from the earlier works of David Hume and had been used by Bentham in *A Fragment on Government* in 1776).

⁷⁹ *Waberley v Cockerel* (1541) 20 Henry VI, Fo 28, Pl 21 in Baker (n 78) 285.

where we have clear and accessible evidence that B has discharged his obligation to A, is there any merit in allowing A to double recover? I tend to think not. A debt is not owed twice.

Given the analysis above, a question remains concerning the potential relevance of economic efficiency and legal certainty to the creation and form of a legal rule. Starting with economic efficiency: the position I take here is that while economic efficiency should not be the exclusive goal of the law, it remains an important and desirable goal for any legal system. In this connection, economic efficiency operates as an important supplementary (or second order) criterion for deciding on one approach over another. To draw an imperfect analogy, economic efficiency has a tiebreaker function (or perhaps more accurately a tiebreaker-like function).⁸⁰ A decision-making analogy might be something like this: A firm agrees to use the ranking of the universities at which two otherwise fairly evenly matched and excellent job candidates applying for the same job studied in order to make a final hiring decision. As applied here, economic analysis can be useful, and of great importance, when choosing between two (or more) different forms that a legal rule may take in the process of translating an abstract and non-consequentialist moral principle into a directly applicable legal rule. For example, the law takes the view that it is morally right to keep a promise, but the doctrine of consideration nonetheless keeps my gratuitous promise to mow my neighbour's lawn out of the courts.⁸¹ Provided a legal rule remains substantively justifiable for non-consequentialist reasons, then one should not ignore law and economics.⁸²

I can make a similar argument with respect to the use of bare appeals to legal certainty in judicial reasoning. It is true that a lack of certainty in the law can be a friend of tyranny. It is a point well made that the law should define its rules in advance and give subjects stable expectations as to how such rules will be deployed.⁸³ It is antithetical to the rule of law for a rule to be made ex post and applied to ex ante facts. As such, I do not wish to be taken as suggesting that certainty is not important to the general law. Rather, I am making the modest claim that certainty, in and of itself, does not exclusively provide a positive justification for a particular legal rule. Let us assume that a state parliament enacts a statute that you

⁸⁰ See further, Dagan and Kreitner (n 65) 575–6. For a critique of the type of reasoning I have deployed, see Robert E Scott and Jody S Kraus, *Contract Law and Theory* (Carolina Academic Press, 5th ed, 2013) 29.

⁸¹ See Hanson, Hanson and Hart (n 74) 324. For examples of balancing interests, see Robert Stevens, *Torts and Rights* (Oxford University Press, 2007) 89 (writing from a rights-based perspective, but nonetheless defending torts that are not actionable per se on the basis of the law needing a floodgate of 'proof of consequential loss' in order to circumscribe a defendant's liability); McBride (n 44) 124 (writing from a natural law perspective).

⁸² While this approach will not lend itself to a logically perfect balance between normativity and efficiency, it nonetheless treats efficiency in the law as an ideal that a legal system (which is otherwise morally justifiable for non-consequentialist reasons) must try to approach even if it can never be perfectly attained.

⁸³ Francis Lieber, *Legal and Political Hermeneutics* (Charles C Little and James Brown, 1839) 88; FA Hayek, *The Road to Serfdom* (Routledge & Kegan Paul Ltd, 1944) 62; Hayek (n 73) 208–9. See also JD Heydon, 'The "Objective" Approach to Statutory Construction' in John Sackar and Thomas Prince (eds), *Heydon: Selected Speeches and Papers* (Federation Press, 2018) 332, 335–6; John Dinwiddy, 'Bentham' in William Twining (ed), *Bentham: Selected Writings of John Dinwiddy* (Stanford University Press, 2004) 54.

pay to me \$100,000 under certain future conditions (for example, if you fail to run 130 km each week for the rest of the calendar year). Even if you happen to be an avid marathon runner, I doubt that you think this rule has much going for it notwithstanding my enthusiasm for the clarity in which the enactment is finally expressed. It is for this reason that there is much wisdom in Williams' observation that 'it is to the interest of legal certainty that, other things being equal, the rules of law should be as clear of application as possible'.⁸⁴ The important point for present purposes, however, is that such a position still raises the ultimate question of whether the 'other things' are indeed equal.

With my caveats on appealing to efficiency and certainty clearly set out, it should be kept in mind that the ambiguity gateway only has a limited effect on the general approach to interpretation. This is because even with the ambiguity gateway, recourse can be had in the first instance to: (i) the text; (ii) the local context within the contractual document (that is, the organising logic and internal structure of the contract — let's call this the 'narrow context'); (iii) the notorious background facts that can be reasonably supposed to be known by both parties; and (iv) commercial common sense. That is, the court already has a fairly expansive set of clues from which to infer what the reasonable recipient of a communication would consider the most probable intention of the author(s).⁸⁵ That is, the ambiguity gateway does not change fundamentally the approach of the court as a matter of the philosophy of language in determining an objective meaning. All it does is remove one set of clues from this process: the relevant background facts and circumstances reasonably known by A and B at the point of entry into the contract (let's call this the 'wide context').

On this view, the interpretive clues available to the court absent ambiguity are likely to be enough to resolve correctly most contractual interpretation disputes without recourse to the wide context. Further, the ambiguity gateway still leaves the parties free to include more contextual information within their contract if they so choose (for example, recitals, definitions and appendices can be used to convert the narrow context into something approaching the wide context).⁸⁶ Accordingly, if the ambiguity gateway does deliver efficiency gains (or can be reformed to deliver efficiency gains), then there is a sound traditional utilitarian basis for the rule. Likewise, a similar argument could be made concerning legal certainty and reducing the potential number of meanings that a contractual text can possibly bear prior to the exercise of judicial power.

While I accept that the efficiency and certainty concerns regarding the resolution of contract disputes are real, there are three brief observations that I wish to make in response. The first is the fact that Australia is an outlier in maintaining the ambiguity gateway throughout the common law world, which should immediately raise questions for those who claim that the gateway is efficient. This is because, as Posner has observed, '[g]lobal consensus (to exaggerate a bit) is further evidence — of course not conclusive — for the optimality of our existing

⁸⁴ Glanville Williams, 'Law and Language — II' (1945) 61 (July) *Law Quarterly Review* 179, 185.

⁸⁵ See also McDougall (n 2) 104: 'contract cases in real life do not often hinge on the distinction between ambiguous and plain language'.

⁸⁶ Schwartz and Scott (n 30) 931, 961–2.

law.⁸⁷ One does not wish to make too much out of Posner's point. But those who make efficiency arguments should consider whether the final level appellate courts in the United Kingdom, Singapore, New Zealand, Hong Kong and Canada (to name a few) are behaving irrationally in creating inefficient rules.⁸⁸

The second observation is that any efficiency benefits derived from the ambiguity gateway may be questioned. As McLauchlan has observed:

The increased costs argument also ignores the reality that many interpretation disputes will be accompanied by alternative claims for rectification of the written contract and possibly also misrepresentation or estoppel, under which evidence of all the negotiations and surrounding circumstances must be received. Accordingly, excluding such evidence for the purpose of interpretation disputes will not have the effect of reducing the length and cost of civil trials.⁸⁹

Likewise, Justice McDougall has noted extra-judicially that in Australia, 'extrinsic evidence is always admissible in the evidentiary sense; that is, courts may always allow its reception ... [i]t is then admissible in the usage sense' if it nevertheless passes the ambiguity gateway.⁹⁰ Indeed, it has become common practice in Australia for cautious trial judges to allow all pre-contractual wide context materials to be adduced as evidence.⁹¹ This is for the principal reason that if the trial judge errs in applying the ambiguity gateway, then the appeal court can interpret the contract in light of the full set of prescribed clues.

Finally, the ambiguity gateway means that efficiency gains are lost as the parties simply tailor a new set of arguments focusing on convincing the court that the text of the contract is ambiguous and that use of the wide context will quell that ambiguity. Indeed, as noted above, this issue could also be used as an appeal point by savvy counsel. Given that large amounts of factual material are nonetheless tendered in contractual disputes and that the ambiguity gateway results in a new species of legal argument centring on ambiguity, it is arguable that the rule is not fully fit for purpose if it is truly concerned with making the resolution of contractual disputes more efficient *in globo*. One partial answer to this argument would be for the High Court of Australia to clarify what is meant by 'ambiguity', by placing a high hurdle for the parties to clear before allowing consideration of the wide context evidence in the interpretive process. In this connection, one commentator has observed that the test for ambiguity set out by the Court of Appeal of New South Wales in *Burns Philp Hardware Ltd v Howard Chia Pty Ltd*,⁹² could serve this function.⁹³ That relatively restrictive approach to ambiguity was set out by Priestley JA (with whom Glass JA agreed) in the following terms:

⁸⁷ Richard A Posner, 'Let Us Never Blame a Contract Breaker' (2009) 107(8) *Michigan Law Review* 1349, 1363.

⁸⁸ I do not wish to make too much of the point, as common law jurisdictions in the US, for example, tend to favour the ambiguity gateway: see above n 30.

⁸⁹ McLauchlan, 'Contractual Interpretation: What Is It About?' (n 3) 37. See also Lindgren (n 4) 166.

⁹⁰ McDougall (n 2) 107.

⁹¹ *McCourt v Cranston* [2012] ANZ Conv R ¶12-006, [24]–[25]. See also Lindgren (n 4) 166.

⁹² *Burns Philp Hardware Ltd v Howard Chia Pty Ltd* (1987) 8 NSWLR 642.

⁹³ Prince, 'Defending Orthodoxy: *Codelfa* and Ambiguity' (n 4) 508.

What I mean by ‘not ambiguous’ for present purposes is not having two or more plausible meanings when the context of the words in the document is taken into account in light of the knowledge any ordinarily intelligent reader of the document would bring to the reading of it.⁹⁴

A similar approach is currently applied in Western Australia, where the court must consider whether there are two or more possible meanings of the impugned provision having regard to: (i) the language of the contract as a whole; (ii) what can be gleaned from the contract itself as to the contractual purpose; and (iii) whether the proffered competing interpretation(s) is/are reasonable.⁹⁵

As noted above,⁹⁶ a more contemporary form of utilitarian reasoning can be deployed to justify the ambiguity gateway (at least in circumstances where sophisticated firms or parties are contracting). In brief, Schwartz and Scott have set out the following three key premises as a justification for the ambiguity gateway in the US.⁹⁷ First, no rule of contractual interpretation can justify devoting infinite resources in order to achieve perfect individualised justice between the parties (that is, something approaching a perfectly accurate interpretation).⁹⁸ It follows that any socially desirable rule of contractual interpretation needs, at some point, to trade-off between, on the one hand, the time taken to draft a contract and litigate contract disputes and, on the other, interpretive accuracy. Second, courts should make this assessment by deferring to actual party preferences and choices regarding interpretation in setting default rules, albeit allowing specific parties to contract out of such rules.⁹⁹ That is, if the majority of contracting parties have an actual unambiguous preference in favour of an ambiguity gateway that fact should create a strong initial presumption in favour of such a rule. This is because the parties themselves are best placed to weigh up the benefits of accuracy, drafting costs and adjudication costs given they directly bear such costs.¹⁰⁰ Third, sophisticated firms and parties have a revealed preference for formal¹⁰¹ rules of interpretation such that the ambiguity gateway (or a similar more textualist approach to interpretation) should be retained, at least in the context of commercial contracting.¹⁰²

⁹⁴ Ibid 657.

⁹⁵ *Technomin Australia* (n 18) 274 [73]–[74]. See also *Siemens Gamesa* (n 18) [99]. In contrast, see the liberal approach in New South Wales adopted in *Newey* (n 14) [89].

⁹⁶ See above nn 65–7 and accompanying text.

⁹⁷ Schwartz and Scott (n 30) 930–5.

⁹⁸ Ibid. Allowing for preference satisfaction enables commercial parties to maximise the profitability of their contractual arrangements (or their contractual ‘surplus’).

⁹⁹ See also Steven Shavell, ‘On the Writing and the Interpretation of Contracts’ (2006) 22(2) *The Journal of Law, Economics & Organisation* 289, 292. Shavell concludes that ‘decisions about the use of [wide context] evidence should be made by the parties, not the courts’: at 307, as reflected in ‘Proposition 6’.

¹⁰⁰ See further Bernheim (n 65) 291–2:

unambiguous choice may nevertheless create a strong presumption concerning well-being the principle of self-determination arguably implies that those involved in governance should judge the impact of interventions on individuals according to the choices those individuals would have made for themselves.

¹⁰¹ On the growth of formalism in US academic writing on contract law, see Avery Wiener Katz, ‘The Economics of Form and Substance in Contract Interpretation’ (2004) 104(2) *Columbia Law Review* 496, 506–12.

¹⁰² Schwartz and Scott (n 30) 930–5, 955–7.

There is not space to do full justice to the arguments presented by Schwartz and Scott. It is a more elegant form of utilitarian reasoning than that typically deployed in Anglo-Australian contract law scholarship.¹⁰³ This is for the simple reason that it eschews the weighing up of costs and benefits in favour of relying on majoritarian choice preferences to ascertain utility (albeit having the benefit of leaving the minority who do not share such preferences a choice or liberty to contract around the proposed ambiguity gateway).¹⁰⁴ I do not intend the use of the word 'simple' in the previous sentence to be taken as disrespectful. Quite the opposite, any reader of Bentham's *Introduction to the Principles of Morals and Legislation*¹⁰⁵ is quickly overwhelmed by the ponderous lists and sub-lists detailing the specific factors that must be weighed up in calculating utility on the classical approach. The use of choice to inform a calculation of utility is an elegant solution that overcomes the difficulty inherent in classical utilitarian balancing exercises. I will endeavour, however, to make a few brief points in response to whether such arguments should, at present, be adopted in Australian jurisprudence.

The first is to make the obvious point that it remains to be seen what actual preferences Australian contracting parties have. More work would need to be undertaken in this regard to sustain a similar argument, but it seems possible that preferences would not change meaningfully between the two sides of the Pacific Ocean. One important difference affecting preferences might be that in the US, limiting admissible evidence and the need to find facts allows a party to apply for summary judgment thereby avoiding a civil jury trial.¹⁰⁶ Concerns regarding keeping a civil jury from affecting¹⁰⁷ the interpretation of a commercial contract or the outcome of a commercial dispute are not a concern in Australia, where such trials are not a relevant feature of the Australian legal landscape.¹⁰⁸ Second, the more

¹⁰³ See further Binmore (n 65) 542–3.

¹⁰⁴ This rationale rests on another economic explanation of the law, game theory or rational choice theory, see Hanson, Hanson and Hart (n 74) 306:

often described as 'the science of strategic thinking,' is a branch of economics concerned with modeling and predicting behavior. Strategic behavior arises when two or more individuals interact and each individual's decision turns on what that individual expects others to do. Game theoretic models have been used to help predict or make sense of everything from chess to childrearing, from evolutionary dynamics to corporate takeovers, and from advertising to arms control.

¹⁰⁵ The best versions of Bentham's works are those produced by the Bentham project at University College London: JH Burns and HLA Hart (eds) *The Collected Works of Jeremy Bentham: An Introduction to the Principles of Morals and Legislation*, (Athlone Press, 1970 ed) 38–50.

¹⁰⁶ Schwartz and Scott (n 30) 932, 943, 960–3; Bix (n 30) 59.

¹⁰⁷ See, eg, Richard A Posner, 'The Law and Economics of Contract Interpretation' (2005) 83(6) *Texas Law Review* 1581, 1603; Edwin W Patterson, 'The Interpretation and Construction of Contracts' (1964) 64(5) *Columbia Law Review* 833, 836–7 (emphasis added):

Even in a case in which a jury is the trier of issues of fact, the interpretation and construction of a written contract is within the exclusive province of the judge. Where the contract is partly oral and partly written, the judge may instruct the jury as to the meaning of the written part, and, with other instructions, leave the remaining issues of interpretation to be determined by the jury. If the jury is directed to bring a general verdict.... It may so doing exercise its views of jury equity and thus impair the reliability of written instruments. *This possibility may account for the reluctance of courts to admit parol evidence and other extrinsic aids to interpretation, and for their adherence to the "plain meaning" of the contract.*

¹⁰⁸ See, eg, *Supreme Court Act 1970* (NSW) s 85. On the decline of the civil jury in New South Wales, see, eg, *Pambula District Hospital v Herriman* (1988) 14 NSWLR 387, 394–7. See also Law Reform

fundamental objection would be to query why the law needs to, or should, reflect majoritarian party preferences and choices. This is because majoritarian preferences will not count for much where such preferences are morally indefensible such that they infringe upon some other normative commitment that the general law should support.¹⁰⁹ At most, it could be said that majoritarian preferences ‘create a strong *presumption* concerning well-being’,¹¹⁰ but that such a presumption is not irrebuttable.

The third point to note is that those who advocate the removal of the ambiguity gateway are not in favour of devoting infinite resources to interpretative disputes. The issue needs to be framed in a way that does not potentially create a false dichotomy whereby the ambiguity gateway is offered as the only potential solution to limit evidence in contractual interpretation disputes. This is because the rules of evidence will still apply to the wide context material, such that the material in question would still need to be relevant and probative in order to be utilised. This point has been emphasised by the Court of Appeal of the Republic of Singapore, which has abolished the ambiguity gateway but nonetheless created specific rules of pleading to ensure that wide contextual evidence is utilised transparently, narrowly and only for legitimate purposes.¹¹¹ Fourth, it is open to question whether the rules of contractual interpretation should apply differing legal standards between sophisticated and unsophisticated parties as a result of the majoritarian preferences of the former but not the latter. Why treat a subset of contracting parties in a differing and unequal (or preferred) way? Further still, at what point will the differing standards apply?¹¹² That is, is the criterion of a sophisticated or a commercial contract precise enough for application? In the local context, it would appear unlikely that the High Court of Australia would be willing to adopt a standard of unequal treatment and create two sets of parallel rules concerning the same activity (that is, abolish the ambiguity gateway for unsophisticated parties, but retain it for sophisticated parties). Of course, the argument I have made does not prevent the High Court maintaining the ambiguity

Commission of Western Australia, *Selection, Eligibility and Exemption of Jurors* (Discussion Paper No 99, September 2009) 11. At the time this Discussion Paper was written, there had been approximately 12 civil jury trials in Western Australia in the preceding four decades. Costs implications are another potential point of difference. Under the American Rule, each party is typically responsible for their own costs: *Alyeska Pipeline Service Co v Wilderness Society*, 421 US 240, 247 (1975). In Australia, on the other hand, party-party costs is the typical order. However, there are exceptions in both jurisdictions, which are not material for present purposes.

¹⁰⁹ See above n 80 and accompanying text. In this connection, one of the strongest points that Schwartz and Scott make is that the ambiguity gateway does not change fundamentally the approach of the court as a matter of the philosophy of language in determining an objective meaning: Schwartz and Scott (n 30) 952, 961. See also the discussion above on this point at n 85 above and accompanying text.

¹¹⁰ Bernheim (n 65) 291 (emphasis added).

¹¹¹ *Sembcorp Marine* (n 28) 225 [73]. See further the text accompanying n 119 below.

¹¹² For a general discussion about not changing default legal rules given the context, see Nicholas Tiverios and Clare McKay, ‘Orthodoxy Lost: The (Ir)relevance of Causation in Quantifying Breach of Trust Claims’ (2016) 90(4) *Australian Law Journal* 231, 240–3; Nicholas Hopkins, ‘The Relevance of Context in Property Law: A Case for Judicial Restraint?’ (2011) 31(2) *Legal Studies* 175. For examples of where the creation of two sets of parallel rules was resisted, see, eg: *Baltic Shipping Co v Dillon* (1993) 176 CLR 344, 365–6 (rejecting the argument that limits on the availability of non-pecuniary loss for breach of contract should be abolished for non-commercial (cf commercial) contracts; *Cavendish Square Holding BV v Makdessi* [2016] AC 1172, 1251 [162] (rejecting the argument that the penalties doctrine should be abolished for commercial (cf non-commercial) contracts).

gateway for all contracting parties — arguments regarding coherence and equal treatment in the law, after all, do not ultimately tell the decision-maker which of two potential forms of a legal rule to select.

As a final point, it should not be overlooked that the context between the parties might nonetheless suggest that a court should give interpretative primacy to textual clues over contextual and purposive clues. For example, think of the common rule that a formal and professionally drafted instrument is to be interpreted more precisely than a communicative act of a lay person.¹¹³ This rule, favouring text over certain aspects of context, is itself a contextual assumption that certain parties generally wish to be taken more literally.¹¹⁴ Sometimes the context may itself point to the parties intending a text to have a narrow or formal meaning. Or, as Morgan has said,

[s]ensitivity to context may actually require the exclusion of broad, contextual interpretation. We have argued above that the detailed drafting of commercial contracts requires a formal, textual interpretive approach. Such contracts are addressed primarily to other lawyers, to be understood in a technical sense (not the 'ordinary understanding' championed by Lord Hoffmann). The relevant context is formalism! The characteristic detailed English drafting style demands textual interpretation.¹¹⁵

On this view, one can arrive at a not dissimilar end point to Schwartz and Scott that allows parties to limit context. This conclusion is reached, however, from a contextualist route — applying the common stock contextual clue that the author of a more formal legal document generally intends it to be read in a formal manner, rather than needing to apply an altogether different legal rule as a result of the majoritarian preferences of sophisticated and unsophisticated parties. If Anglo-Australian courts are willing to apply a common stock contextual clue or assumption that parties behaving formally intend to be taken more literally, then there appears to be no reason in principle why the parties cannot expressly stipulate such an intention for themselves (as I outline in this article, the argument for context in contractual interpretation is, after all, based itself on intentionality). Put another way, the parties' express intentions regarding how their language is to be interpreted¹¹⁶ should matter just as much as their assumed intentions.

¹¹³ For a simple example of this common principle in action, see *Thorney Park Golf* (n 43) [24] (McCombe LJ).

¹¹⁴ A point made in Morgan (n 63) 233.

¹¹⁵ *Ibid.*

¹¹⁶ Making a similar point, but not from a contextualist perspective, see Katz (n 101) 514, 521–2. See also Posner, who notes that arguments in favour of a wide approach to context often do not consider that contracting parties can also have intentions regarding how a contract (ie, the manifestation of their intentions) is to be interpreted: Eric A Posner 'The Parol Evidence Rule, The Plain Meaning Rule, and The Principles of Contractual Interpretation' (1998) 146(2) *University of Pennsylvania Law Review* 533, 569–71.

IV Other Options – If Not an Ambiguity Gateway then What?

It is important to observe that the ambiguity gateway is not the only option when it comes to attempts to make the resolution of contractual disputes more efficient. There are two other obvious solutions — although it should be conceded that such approaches could nonetheless work in concert with an ambiguity gateway or a revised version of that principle. First, the principles concerning active case management and costs orders could inform more effective mechanisms for improving the dispute resolution process.¹¹⁷ In this connection, Arden LJ observed the potential relevance of case management principles in *Static Control Components (Europe) Ltd v Egan*:

When the principles in the *ICS* case were first enunciated, there were fears that the courts would on simple questions of the construction of deeds and documents be inundated with background material. Lord Hoffmann recognised this risk by emphasising in *BCCI v Ali* [2002] 1 AC 251 at 269 that his reference to ‘absolutely anything’ in his second proposition was to anything that a reasonable man would have regarded as relevant. Speaking for myself, I am not aware that the fears expressed as to the opening of floodgates have been realised. The powers of case management in the CPR could obviously be used to keep evidence within its proper bounds. The important point is that the principles in the *ICS* case lead to a more principled and fairer result by focussing on the meaning which the relevant background objectively assessed indicates that the parties intended.¹¹⁸

Second, like the use of active case management, the rules of pleading in contractual disputes can seek to limit the breadth of the more contextual *Investors Compensation Scheme* principles and the impact of those principles on the efficiency of contractual disputes. In Singapore, for example, the creation of new rules of pleading have sought to achieve this by limiting the need for a judge to wade through the potentially voluminous thicket of pre-contractual evidence in order to find the contextual ‘needle in a haystack’. Rather, the burden has been placed on the party bringing the haystack into court to point to the needle. Menon CJ set out these principles of pleading in *Sembcorp Marine Ltd v PPL Holdings Pte Ltd*:

to buttress the evidentiary qualifications to the contextual [*Investors Compensation Scheme*] approach to the construction of a contract, the imposition of four requirements of civil procedure are, in our view, timely and essential:

- (a) first, parties who contend that the factual matrix is relevant to the construction of the contract must plead with specificity each fact of the

¹¹⁷ See McLauchlan, ‘Contractual Interpretation: What Is It About?’ (n 3) 11. Although note Schwartz and Scott, who argue that judges do not bear the costs of litigation themselves and may thus tend to prefer individualised justice and contextual interpretive accuracy in comparison to the parties (if this is true, then such a preference for contextual material could colour how judges ultimately use case management principles): see Schwartz and Scott (n 30) 943.

¹¹⁸ *Static Control Components (Europe) Ltd v Egan* [2004] 2 Lloyd’s Rep 429, 435–6 [29]. See also Donald Nicholls, ‘My Kingdom for a Horse: The Meaning of Words’ (2005) 121 (October) *Law Quarterly Review* 577, 588.

factual matrix that they wish to rely on in support of their construction of the contract;

- (b) second, the factual circumstances in which the facts in (a) were known to both or all the relevant parties must also be pleaded with sufficient particularity;
- (c) third, parties should in their pleadings specify the effect which such facts will have on their contended construction; and
- (d) fourth, the obligation of parties to disclose evidence would be limited by the extent to which the evidence are relevant to the facts pleaded in (a) and (b).¹¹⁹

Ultimately, any sound utilitarian analysis requires the analyst to get her sums right.¹²⁰ It is difficult to state with any certainty which approach best maximises efficiency gains (at least in the traditional sense) without a detailed empirical analysis of the costs of, and benefits to, the efficient resolution of contractual disputes associated with the application of the ambiguity gateway, the potential use of case management principles and potential changes to rules of pleading. Rather, the goal here has been modest: to raise some tenable alternatives to the ambiguity gateway principle in order to make the resolution of contractual disputes more efficient given the benefits that the *Investors Compensation Scheme* approach otherwise provides to the interpretive process.

V Conclusion

In *Sirius International Insurance Co v FAI General Insurance Ltd*, Lord Steyn quoted the words of the famous Christian apologist William Paley: ‘the tyrant Temures promised the garrison of Sebastia that no blood would be shed if they surrendered to him. They surrendered to him. He shed no blood. He buried them all alive’.¹²¹ The principal reason why the reader of this quotation knows that Temures committed an injustice is that we all intuitively know the difference between, on the one hand, the sentence meaning of an utterance and what, on the other hand, a reasonable recipient of an utterance would believe the speaker meant. As I have illustrated in this article, the objective intention that the court searches for in a contractual interpretation dispute is distinct from the sentence meaning of a text. If legal interpretation only cared about sentence meaning, then the task of the court would be mercifully narrow: to decode the literal meaning of a text. This is not the modern law of interpretation. It is a basal principle that a contract is to be interpreted by the reasonable recipient of the communication read contextually and purposively. The surrounding circumstances form part of that context such that the ambiguity gateway deprives the court of otherwise probative and relevant evidence in the interpretative process. While the response to this argument is that the ambiguity gateway assists in the efficient resolution of disputes, it is incumbent on those

¹¹⁹ *Sembcorp Marine* (n 28) 225 [73].

¹²⁰ See generally Burns and Hart (n 105) ch 4.

¹²¹ *Sirius International Insurance Co v FAI General Insurance Ltd* [2005] 1 WLR 3251, 3258, quoting (with minor formatting differences) William Paley, *The Works of William Paley* (Longman and Co, 1838) vol III, 60. Also cited in Johan Steyn, ‘The Intractable Problem of the Interpretation of Legal Texts’ (2003) 25(1) *Sydney Law Review* 5, 7.

making this utilitarian claim to get their sums right and to justify their conclusions as to the desirability of efficiency as an end goal of the law. I am willing, at least at present, to remain open minded. If such justifications remain wanting, then the ambiguity gateway should be abolished.

Before the High Court

Liability for Workplace Psychiatric Injury and Vicarious Trauma: *Kozarov v Victoria*

Kylie Burns*

Abstract

Work-related psychiatric injury claims are frequent and costly. Workplace psychosocial hazards increase the risk of prolonged workplace stress, which can lead to physical and psychological injury. This article considers the forthcoming appeal to the High Court of Australia from the Victorian Court of Appeal decision in *Victoria v Kozarov*, which concerns a psychiatric injury because of vicarious trauma in the workplace. The appeal raises important issues about the application of principles enunciated by the High Court in *Koehler v Cerebos (Australia) Ltd*. In *Kozarov*, the High Court will consider the test of reasonable foreseeability in a context where an employer had actual knowledge of the risk of psychological injury to all employees as a result of vicarious trauma. Additionally, the appeal raises issues about: the role of employment contracts in determining negligence; the emphasis to be given to issues of privacy and autonomy in defining the scope of an employer's duty of care; inferential factual reasoning in causation; and the interaction between an employer's obligation to enforce a safe system of work and counterfactual causation.

Please cite as:

Kylie Burns, 'Liability for Workplace Psychiatric Injury and Vicarious Trauma: *Kozarov v Victoria*' (2021) 43(4) *Sydney Law Review* 575.



This work is licensed under a Creative Commons Attribution-NonDerivatives 4.0 International Licence (CC BY-ND 4.0).

As an open access journal, unmodified content is free to use with proper attribution. Please email sydneylawreview@sydney.edu.au for permission and/or queries.

© 2021 Sydney Law Review and author. ISSN: 1444-9528

* Associate Professor, Griffith Law School, Griffith University, Nathan, Queensland, Australia. Email: k.burns@griffith.edu.au; ORCID iD: <https://orcid.org/0000-0001-9303-2252>.

I Introduction

Work-related psychiatric injury claims are frequent and costly.¹ Individual vulnerability of employees is not the only cause of workplace psychiatric injury. Workplace psychosocial hazards increase the risk of prolonged workplace stress, which can lead to physical and psychological injury.² Hazards include long work hours, heavy workloads and job demands, hazardous or high pressure environments, distressed or aggressive clients, exposure to violence and trauma, poor workplace support, and workplace bullying, harassment and sexual assault.³ There are particular risks of vicarious trauma to employees such as lawyers, child protection workers, police, ambulance officers, journalists, and forensic scientists who witness or are exposed to traumas experienced by others.⁴ The Productivity Commission's Mental Health Inquiry,⁵ the Australian Human Rights Commission's National Inquiry into Sexual Harassment in Australian Workplaces⁶ and other recent inquiries⁷ have identified the need for employers to take greater steps to prevent and adequately respond to workplace psychiatric injuries.

The High Court of Australia has granted special leave⁸ to appeal the Victorian Court of Appeal decision in *Victoria v Kozarov*,⁹ a case concerning psychiatric injury arising from vicarious trauma in the workplace. The case raises important issues about the application of principles enunciated by the High Court's 2005 decision in *Koehler v Cerebos (Australia) Ltd*.¹⁰ In *Koehler*, which concerned a workplace psychiatric injury caused by overwork, the High Court held that while it may be 'general knowledge that some recognisable psychiatric illnesses may be triggered by stress', it was a 'further and much larger step' to expect that all employers must recognise the risk of psychiatric injury to all employees from stress at work.¹¹ While employers owed a general duty of care to employees to provide a safe system of work, foreseeability of psychiatric harm to particular employees may depend on factors including the nature and extent of the work, and explicit or implicit

¹ Productivity Commission, *Mental Health: Inquiry Report* (Report No 95, June 2020) vol 2, 308–12 <<https://www.pc.gov.au/inquiries/completed/mental-health/report/mental-health-volume2.pdf>>.

² Safe Work Australia, *Work-Related Psychological Health and Safety: A Systematic Approach to Meeting Your Duties* (National Guidance Material, January 2019) 9–12 <https://www.safeworkaustralia.gov.au/system/files/documents/1911/work-related_psychological_health_and_safety_a_systematic_approach_to_meeting_your_duties.pdf>.

³ *Ibid*.

⁴ *Ibid* 12.

⁵ Productivity Commission (n 1).

⁶ Australian Human Rights Commission, *Respect@Work: Report of the National Inquiry into Sexual Harassment in Australian Workplaces* (2020) Section 5 <<https://humanrights.gov.au/our-work/sex-discrimination/publications/respectwork-sexual-harassment-national-inquiry-report-2020>>.

⁷ See, eg Marie Boland, *Review of the Model Work Health and Safety Laws: Final Report* (December 2018) 30–5 <https://www.safeworkaustralia.gov.au/system/files/documents/1902/review_of_the_model_whs_laws_final_report_0.pdf> ('*Boland Review Report*').

⁸ Transcript of Proceedings, *Kozarov v Victoria* [2021] HCATrans 101, 804.

⁹ *Victoria v Kozarov* (2020) 301 IR 446 ('*Kozarov* (VSCA)') which allowed an appeal by the State of Victoria against the decision in *Kozarov v Victoria* (2020) 294 IR 1 ('*Kozarov* (VSC)').

¹⁰ *Koehler v Cerebos (Australia) Ltd* (2005) 222 CLR 44 ('*Koehler*').

¹¹ *Ibid* 57 [34] (McHugh, Gummow, Hayne and Heydon JJ). See also 64–5 [54]–[57] (Callinan J).

signs from the employee.¹² The scope of duty could be constrained by: the employment contract; equitable obligations; coherence with relevant legislative frameworks; and by the employee's own agreement to carry out duties.¹³ Since *Koehler*, additional barriers to recovery in negligence have emerged, reliant on policy considerations of employee privacy, dignity and autonomy.¹⁴ In *Kozarov*, the High Court will consider the test of foreseeability in a context where (unlike *Koehler*) an employer had actual knowledge of the risk of psychological injury to all employees as a result of vicarious trauma and had reflected this in its own policies. Additionally, the case raises issues around: the role of employment contracts in defining the scope of duty of care and negligence principles;¹⁵ the emphasis to be given to issues of privacy and autonomy in defining scope of duty;¹⁶ inferential factual reasoning in causation; and the interaction between an employer's obligation to enforce a safe system of work¹⁷ and counterfactual causation. Broader questions arise about how 'common sense' social facts about the nature and causation of workplace psychiatric injuries affect determinations about scope of duty, breach and causation.¹⁸

II The Facts and Procedural History

The appellant, Zagi Kozarov ('Kozarov'), was employed by the Victorian Office of Public Prosecutions ('OPP') as a solicitor in their Specialist Sexual Offences Unit ('SSOU') between June 2009 and April 2012.¹⁹ She was exposed to sexual offences cases involving children including 'graphic images' of rapes, assaults and child pornography and prepared children for evidence and cross examination.²⁰ She dealt with the 'extremely' distressing aftermath of verdicts including suicidal complainants.²¹ She frequently worked excessive hours.²² She developed work-

¹² Ibid 54 [23]–[24], citing Hale LJ in the English Court of Appeal decision in *Sutherland v Hatton* [2002] 2 All ER 1, 14 [25]–[28]. The High Court of Australia in *Koehler* (ibid [24]) suggested there may be other matters, not explored in argument, that might make the risk of psychiatric injury reasonably foreseeable and that would require 'deeper knowledge' about causes of psychiatric injury than general knowledge.

¹³ *Koehler* (n 10) 53–7 [21]–[25], [28]–[31].

¹⁴ See, eg, *Hegarty v Queensland Ambulance Service* (2007) Aust Torts Reports ¶81-919 ('*Hegarty*').

¹⁵ While *Koehler* raised the content of employment contracts as a relevant factor in determining the scope of duty of care (with other factors), this aspect was not fully evaluated by the High Court which noted that this would have required 'much closer attention to the content of the contractual relationship than was given in the evidence and arguments in the courts below': *Koehler* (n 10) 58 [40]. See also [38]–[39].

¹⁶ This issue has not yet been fully explored in the High Court following the refusal of special leave to appeal *Hegarty* (n 14): see *Hegarty v Queensland Ambulance Service* [2008] HCATrans 121.

¹⁷ *McLean v Tedman* (1984) 155 CLR 306, 313 ('*McLean*').

¹⁸ For discussion of risks in judicial use of common-sense social facts, see Kylie Burns, 'It's not just Policy: The Role of Social Facts in Judicial Reasoning in Negligence Cases' (2013) 21(2) *Torts Law Journal* 73; Kylie Burns, 'Judges, "Common Sense" and Judicial Cognition' (2016) 25(3) *Griffith Law Review* 319.

¹⁹ *Kozarov (VSCA)* (n 9) 448 [1].

²⁰ Ibid 452 [18]–[21].

²¹ Ibid 452 [21].

²² Ibid 452–3 [22]–[23]. Evidence from fellow employees suggested heavy workload was common: at 453 [23].

related chronic post-traumatic stress disorder ('PTSD') and a major depressive disorder in 2011–12.²³ There was extensive evidence that the OPP was aware of the significant risks of vicarious trauma to employee health.²⁴ The SSOU manual described management responsibility for risk identification and management and required compliance with Occupational Health and Safety ('OHS') legislation.²⁵ A Vicarious Trauma Policy ('VT Policy') published in 2008 acknowledged the impacts of vicarious trauma on SSOU staff; outlined preventative strategies such as avoiding excessive hours and excessive workload; and specified management options including 'rotations within the OPP, counselling, debriefing, the relocating of files, specific "time outs" and the provision of assistance to staff members'.²⁶ Despite this, Kozarov's managers were 'unaware of these documents'²⁷ and SSOU staff and management knowledge of vicarious trauma was 'desultory'.²⁸

In March 2011, SSOU staff held a staff meeting and wrote a memo on staff-wellbeing to management complaining of work overload and outlining stress-related symptoms.²⁹ In April 2011, a psychologist engaged by the SSOU held a session where staff including Kozarov discussed the impact of work on their private lives.³⁰ In June 2011, Kozarov was allocated an additional file despite her complaints she was already overloaded and working long hours and weekends.³¹ In August 2011, she became unwell at work and was on sick leave for several weeks.³² Upon return to work in late August 2011, Kozarov was involved in a conflict with her manager when he incorrectly assumed she had come into work late. This was followed by a series of lengthy distressed emails from her to her manager.³³ In late October 2011, she attended a meeting with management and raised concerns by junior staff about the confronting nature of SSOU work.³⁴ From September to December 2011, Kozarov continued to experience intense workload.³⁵ She accepted a promotion to a permanent higher-level position before taking recreation leave and long service leave.³⁶ On 9 February 2012, when she was to return from leave, she emailed her managers advising them of the severe effects that exposure to SSOU matters had had on her psychological health and requested an internal transfer.³⁷ Following

²³ Ibid 448 [1].

²⁴ Ibid 453–4 [24]–[25]. Between 2007 and 2009 this included management emails, staff training sessions, an evaluation by a psychologist, a briefing paper about loss of staff, and memo from the SSOU principal solicitor about staff wellbeing, stress and vicarious trauma.

²⁵ Ibid 451 [10].

²⁶ Ibid 451 [12].

²⁷ Ibid 454 [25].

²⁸ Ibid, quoting *Kozarov (VSC)* (n 9) 37 [149].

²⁹ *Kozarov (VSCA)* (n 9) [30]–[34]. A copy of the memo was later sent by the SSOU manager to senior OPP staff with a business case for additional staff: at 457 [36].

³⁰ Ibid 456 [35].

³¹ Ibid 457 [39]–[40].

³² Ibid 458 [41]–[43].

³³ Ibid 458–61 [45]–[48].

³⁴ Ibid 461 [52].

³⁵ Ibid 461 [49].

³⁶ Ibid 461 [53]. She had applied for the promotion on 28 August 2011 while on sick leave: at 458 [44].

³⁷ Ibid 462–3 [54]–[56].

unsuccessful attempts to return her to work in the OPP from February to April 2012, her employment was terminated.³⁸

At trial, Jane Dixon J held that the OPP had breached their duty of care, which caused Kozarov psychiatric injury. Dixon J concluded that internal documents and policies showed the OPP knew of the risks to SSOU staff from burnout, work stress and vicarious trauma.³⁹ The OPP was also aware of specific risks to Kozarov through a combination of ‘evident signs’ of her declining mental health, culminating with the conflict and emails when she returned from sick leave late August 2011 (the ‘sentinel event’⁴⁰).⁴¹ By this stage, her psychiatric injury was reasonably foreseeable to the OPP. Dixon J found that a reasonable employer would have taken a range of steps to ensure a safe system of work including

an active OH&S framework; more intensive training for management and staff regarding the risks to staff posed by vicarious trauma and PTSD; welfare checks and the offer of referral for a work-related or occupational screening, in response to staff showing heightened risk; and, a flexible approach to work allocation, especially where required in response to screening, including the option of temporary or permanent rotation from the SSOU where appropriate.⁴²

‘Poor handling’ of the return-to-work process, which continued to expose Kozarov to sexual offences, was a ‘continuing breach of duty’ that added to the severity of her injury.⁴³

At trial, Kozarov made an alternative contractual claim.⁴⁴ Unlike *Koehler* case, Kozarov’s employment contract had extensive provisions about reasonable workload and OHS obligations and referred to OPP policies including the SSOU Manual. Dixon J found that ‘[u]nlike *Koehler*, the obligations between the parties under the employment contract strengthen[ed], rather than weaken[ed], [Kozarov’s] claim’ and were ‘highly relevant’ to the negligence claim.⁴⁵ Contractual breaches did not, however, raise separate legal issues to the negligence claim and did not result in a separate assessment of damages.⁴⁶ Kozarov also relied initially on a claim of breach of statutory duty arising from breach of the *Occupational Health and Safety Act 2004* (Vic) and associated regulations. ‘[M]inimal attention’ was paid to this claim at trial and Dixon J held that the onus of proof was not satisfied.⁴⁷ Her Honour also found that damages should not be reduced for any alleged contributory negligence.⁴⁸

³⁸ Ibid 463 [58].

³⁹ Ibid 463 [61].

⁴⁰ See *ibid* 450 [7], citing *Kozarov* (VSC) (n 9) 134 [598], 136–7 [609].

⁴¹ *Kozarov* (VSCA) (n 9) 463–5 [62]–[66].

⁴² Ibid 465 [67], quoting *Kozarov* (VSC) (n 9) 155–6 [702].

⁴³ *Kozarov* (VSC) (n 9) 166 [750]. See also 165–6 [734]–[749].

⁴⁴ Ibid 166–9 [751]–[767]. The contract referred to the Victorian Public Service Agreement 2006 (2009 Extended and Varied Version) <<https://www.fwc.gov.au/documents/awards/tracee/agreements/pdf/ag847284.pdf>> and the *Public Service Administration Act 2004* (Vic).

⁴⁵ *Kozarov* (VSC) (n 9) 169 [767].

⁴⁶ Ibid 169 [766].

⁴⁷ Ibid 169 [768].

⁴⁸ Ibid 170–1 [769]–[776].

The OPP appealed to the Court of Appeal of Victoria on two grounds. The first ground, relating to duty of care and breach of duty, alleged error in finding that a sentinel event had occurred and that there were evident signs of Kozarov's psychological injury by the end of August 2011, which warranted response from the OPP. The second ground, relating to causation, alleged error in finding that if there had been no breach, Kozarov's injury would have been avoided. The Court of Appeal dismissed the first ground, but allowed the appeal in relation to causation. It was not satisfied that even if Kozarov had been made aware of her PTSD and continuing risk to her mental health by the end of August 2011, she would have agreed to be transferred or could have been compelled to work in another work unit.⁴⁹ The Court placed particular emphasis on Kozarov's statements in the sentinel event emails defending her work record and indicating her passion for work, and on her decision to later pursue and accept promotion.

Kozarov's first ground of appeal in the High Court relates to causation. Kozarov submits that the Court of Appeal erred by overturning an inference drawn by Dixon J at trial that if the OPP had discharged their duty of care so that Kozarov was aware of her injury, she would have been 'given time away from [her] confronting work, and thereby not suffered [her] injury'.⁵⁰ Kozarov's second ground of appeal relates to coherency between scope of duty and causation. Kozarov submits that the Court of Appeal erred 'in failing to consider the nature and content of the [OPP's] duty of care include[ing] ... a duty to maintain and enforce a safe system of work', when determining counterfactual causation.⁵¹ The OPP raises a further issue by notice of contention concerning duty of care and breach, arguing that the Court of Appeal erred in finding that the OPP was on notice of risk to Kozarov's health by the end of August 2011 so as to require reasonable steps to be taken by the OPP in response.⁵²

III Critique and Commentary

A *Duty of Care, Reasonable Foreseeability and 'Evident Signs'*

The Court of Appeal found that the *Koehler* principles required that in order to engage a duty of care to a particular employee concerning psychiatric injury, 'evident signs' of particular risk of psychiatric illness to that particular employee are required.⁵³ Neither the OPP nor Kozarov challenge the *Koehler* principles in the High Court. There is, rather, a dispute about when evident signs of illness could be reasonably recognised by the OPP. At trial and in the Court of Appeal, it was held that there were cumulative evident signs that Kozarov was at risk of psychiatric

⁴⁹ *Kozarov (VSCA)* (n 9) 478–9 [106]–[110].

⁵⁰ Zagi Kozarov, 'Appellant's Submissions', Submission in *Kozarov v Victoria*, Case No M36/2021, 9 July 2021 [3] <https://cdn.hcourt.gov.au/assets/cases/06-Melbourne/m36-2021/Kozarov-Vic_App.pdf> ('Kozarov's Submissions').

⁵¹ *Ibid.*

⁵² *Ibid.*

⁵³ *Kozarov (VSCA)* (n 9) 466 [69].

injury by the end of August 2011.⁵⁴ The OPP argues that a duty of care ‘was not engaged’ until February 2012 when Kozarov notified she was unable to return to work and requested reassignment.⁵⁵ Her injury was not reasonably foreseeable until that late stage. Factually interpreting what ‘evident’ signs of psychiatric injury are and when (in retrospect) those signs would have been obvious enough to a reasonable employer is a fraught and difficult exercise. The *Koehler* principles clearly indicate that the nature of the work itself is an important factor in determining reasonable foreseeability. The OPP’s explicit knowledge of the potential signs and risks of vicarious trauma due to SSOU work are an important framework for interpreting the ‘signs’ of Kozarov’s illness. Given this framework, the Court of Appeal and trial judge’s finding that evident signs were present in August 2011 seems preferable.

There are, moreover, broader issues that may emerge in the *Kozarov* appeal about whether there is a need to clarify or re-evaluate the *Koehler* principles. Is it still appropriate to maintain that the risks of workplace psychiatric injury are not generally reasonably foreseeable to all employers such that focus must be on both the nature and extent of the work done and explicit or implicit signs from each particular employee to trigger a duty of care? This common sense proposition was controversial following *Koehler*.⁵⁶ In 2021, it is incongruous to suggest Australian employers should not be aware of the risks to employee health from workplace psychosocial hazards.⁵⁷ Employers are required by OHS legislation to identify and manage (proactively and reactively) the risks of psychiatric injury in the same way as the risks of physical injury.⁵⁸ There should be coherence between negligence law and OHS legislation. Where an employer foresees and has knowledge of significant particular risks of psychiatric injury to all their employees due to the nature of the workplace, does the test of reasonable foreseeability also require explicit or implicit

⁵⁴ Ibid 466–72 [70]–[84].

⁵⁵ Ibid 463 [57].

⁵⁶ See, eg, Rima Hor, ‘Psychiatric Illness in the Workplace: The Implications of *Koehler v Cerebos*’ (2005) 27(3) *Sydney Law Review* 557.

⁵⁷ The Victorian WorkSafe ‘Stresswise Policy’ was referred to by an expert OHS witness during the trial in *Kozarov v Kozarov* (VSC) (n 9) 143 [634]. Current Victorian resources include: WorkSafe Victoria, Work-Related Stress (Website, 30 August 2021) <<https://www.worksafe.vic.gov.au/work-related-stress>>; WorkSafe Victoria, *A Guide for Employers: Preventing and Managing Work-Related Stress* (February 2021) <<https://www.worksafe.vic.gov.au/resources/preventing-and-managing-work-related-stress-guide-employers>>. Chapter 17 of the *Guide for Employers* deals with trauma exposure injuries. See also the reports referred to above in nn 1–2, 6–7.

⁵⁸ Safe Work Australia (n 2). The *Boland Review Report* recommended amendment of the work health and safety regulations to reiterate the obligations of employers to identify, manage and control psychological workplace injury: *Boland Review Report* (n 7) 30–5. Changes are in the process of being introduced across Australia: see, eg, SafeWork NSW, NSW Government, *Code of Practice: Managing Psychosocial Hazards at Work* (May 2021) <https://www.safework.nsw.gov.au/__data/assets/pdf_file/0004/983353/Code-of-Practice_Managing-psychosocial-hazards.pdf>. The Victorian Government has announced it will also implement new regulations to provide further guidance to employers to protect workers from mental injury: Ingrid Stitt (Minister for Workplace Safety), ‘Keeping Workers Safe from Psychological Harm’ (Media Release, Victoria State Government, 17 May 2021) <<https://www.premier.vic.gov.au/keeping-workers-safe-psychological-harm>>.

signs of injury from a particular employee?⁵⁹ In *Kozarov*, the OPP was actually aware of the significant risk of psychiatric injury to SSOU employees from vicarious trauma. The OPP's VT Policy (which was not properly implemented) specifically identified symptoms and strategies to ameliorate that risk. It seems logically incoherent that a duty of care to take care to prevent an injury, where the high risk of injury is already known by an employer, may arise only after the injury has occurred such as to manifest 'evident' signs.⁶⁰ Such a duty is meaningless, without content and comes too late. For example, requirements to have a safe system of work that includes preventative proactive steps to detect and respond to the risk of psychiatric injury may apply before the duty arises.⁶¹ If a duty of care only arises after an employee is already so ill that evident signs of psychiatric illness have manifested, it also becomes more difficult to satisfy causation. It may be difficult for an employee to show the injury could have been prevented by 'reasonable' employer precautions that only arrive late and reactively.

B Breach, Early Intervention and Privacy

It is significant that the *Kozarov* appeal will consider the role of employee privacy and autonomy in determining scope and breach of an employer's duty of care.⁶² At trial, the OPP was found to have breached their duty of care by failing to provide a safe system of work, which included a failure to make a welfare inquiry or offer occupational screening or follow up measures. The OPP argues in the High Court that this impermissibly formulates an 'unrealistic duty to intrude into an employee's mental well-being'.⁶³ This can be traced to Keane JA's judgment in *Hegarty v Queensland Ambulance Service*, which suggested that an obligation on employers to follow up and intervene may impinge on the dignity of individuals and may intrude on their private life, autonomy and privacy.⁶⁴ The judgment suggested, by way of 'social, economic and legal context', that employees may be 'deeply resentful' if

⁵⁹ There may be an argument that *Koehler*, which concerned work overload and general workplace stress, left this question unresolved and recognised there may be other factors that may make an injury reasonably foreseeable: see *Koehler* (n 10) 54–5 [24].

⁶⁰ See, eg, *Melville v Home Office* (a case concerning a prison officer who had to attend prisoner suicides) heard with other cases in the English Court of Appeal in *Hartman v South Essex Mental Health and Community Care NHS Trust*, where it was held that the employer's knowledge of the risk of significant trauma to its employees evidenced by its own policies was sufficient to satisfy foreseeability without further signs of vulnerability from the employee: *Hartman v South Essex Mental Health and Community Care NHS Trust* [2005] EWCA Civ 6. See also discussion by McColl JA in *New South Wales v Briggs* in relation to the foreseeability of injury to employees regularly exposed to trauma due to their occupation: *New South Wales v Briggs* (2016) 95 NSWLR 467, 471–7 [2]–[30] ('*Briggs*').

⁶¹ This was the case in *Kozarov*, where one of the breaches included failures to have a proactive OHS system including staff training to allow early recognition of and response to symptoms: *Kozarov* (VSC) (n 9) 144 [643], 151 [676]–[677].

⁶² See above n 16 and accompanying text.

⁶³ State of Victoria, 'Submissions of the Respondent', Submission in *Kozarov v Victoria*, Case No M36/2021, 6 August 2021, [55] <https://cdn.hcourt.gov.au/assets/cases/06-Melbourne/m36-2021/Kozarov-Vic_Res.pdf> ('Victoria's Submissions'). See also [24], [56]. This was not an issue considered in the Court of Appeal.

⁶⁴ *Hegarty* (n 14) [41]. See also *Briggs* (n 60) 497 [126]–[127] (Leeming JA).

their employers intervened when concerned an employee may be showing signs of mental illness.⁶⁵ Employees may consider this to be ‘a gross impertinence’ and such interventions may lead to employee grievances, industrial issues and potentially defamation actions.⁶⁶ These statements about potential consequence appear to be assumptions of judicial common sense or of judicial notice.⁶⁷

At trial, Dixon J was ‘mindful’⁶⁸ of Keane JA’s comments in *Hegarty* and accepted the ‘the need to avoid unnecessarily impinging on the personal autonomy of professional staff’.⁶⁹ However her Honour found that an employer could not be immune from making proper inquiries about staff welfare when the circumstances warranted and ‘a system of work that openly acknowledged the risks attached to the work and offered welfare inquiries or screening’ would uphold employee dignity and prevent workplace injury.⁷⁰ A welfare inquiry and offer of workplace screening to Kozarov was appropriate given inherent and known risks in the SSOU, general signs of distress in the SSOU and evident signs of Kozarov’s own distress.⁷¹ It was consistent with the OPP’s own VT Policy which indicated that ‘a personal approach needed to be made if vicarious trauma was suspected’⁷² and with the contractual obligations of the OPP as a public sector employer.⁷³ In high risk environments, it was insufficient to place an onus on individual staff to initiate conversations about health concerns.⁷⁴

The approach of Dixon J is the preferable approach. *Hegarty* is distinguishable from *Kozarov*. Both the trial and appellate court in *Kozarov* held that there were clear signs that Kozarov was at risk of psychiatric harm warranting intervention. In *Hegarty*, there were no clear signs that the plaintiff was at risk of any psychiatric harm and it was in that context that welfare inquiries or intervention were held to be unwarranted and overly intrusive.⁷⁵ Policy concerns such as personal dignity, privacy and autonomy may be relevant considerations in some cases in determining the scope of a duty of care or breach. However, ‘[c]are must be taken to ensure that solicitude for an employee’s privacy does not overwhelm those other considerations that give rise to a meaningful duty of care to avoid injury’.⁷⁶ Coherence favours negligence principles that are consistent with legislative OHS requirements that suggest early intervention approaches.⁷⁷ Early intervention conversations can be evidence-based, carried out compassionately with privacy and

⁶⁵ *Hegarty* (n 14) [46].

⁶⁶ *Ibid.*

⁶⁷ See Burns articles cited above in n 18.

⁶⁸ *Kozarov* (VSC) (n 9) 133 [593].

⁶⁹ *Ibid* 150 [670].

⁷⁰ *Ibid* 151 [676].

⁷¹ *Ibid* 150 [670]. See also the discussion of previous cases at 136–9 [608]–[619] where there was found to be absence of notice, which made them distinguishable from *Kozarov*.

⁷² *Ibid* 150 [670].

⁷³ *Ibid* 138–9 [617]–[618].

⁷⁴ *Ibid* 151 [677].

⁷⁵ *Hegarty* (n 14) [97]–[102].

⁷⁶ *The Age Co Ltd v YZ* (2019) 60 VR 189, 216 [131].

⁷⁷ See above n 57.

confidentiality and can support the worker to continue or return to work.⁷⁸ Courts should also be cautious about making personal observations or determining what social interests dictate in workplace psychiatric injury cases based on judicial perception.⁷⁹ Judicial common-sense social fact assumptions about how employees are likely to behave or respond in the workplace may be ill-founded.⁸⁰ Over-focus on ‘privacy’ may foster non-intervention by employers and lack of preventative measures inconsistent with modern workplace management. This may reinforce stigma about mental illness, which increases rates and effects of injury.⁸¹

C Causation of Injury

The critical causation question in the High Court is whether the exacerbation of Kozarov’s injury could have been prevented after August 2011. This raises difficult issues including the role of factual inferences in determining causation and the relevance of employment contracts and an employer’s obligation to enforce a safe system of work to counterfactual causation. At trial, Dixon J was satisfied that if a welfare inquiry had been made by the OPP, Kozarov would have taken up an offer of psychological screening, which would have ultimately resulted in ‘altering work allocation, or arranging time out, or rotation to another role, if required’.⁸² Dixon J found that it was likely Kozarov would have cooperated with change in work allocation away from the SSOU if she had been informed of the rationale.⁸³ The Court of Appeal found that the only response that would have prevented the exacerbation of Kozarov’s injury after the end of August 2011 was rotation, given it was likely she would have been diagnosed with PTSD at that time.⁸⁴ The Court held that there was no suggestion that she could have been compelled to accept a rotation and made the important conclusion (in a single sentence without detailed analysis) that the employment contract would have precluded the OPP rotating Kozarov to another role.⁸⁵ To satisfy causation, Kozarov was therefore required to show that she would have voluntarily accepted that rotation if offered a rotation of role away from the SSOU at the end of August 2011. At trial, Dixon J found that her requests in February 2012 when she advised she was ill and requested a transfer away from the SSOU demonstrated that she would have accepted a rotation in August 2011 if she

⁷⁸ Safe Work Australia (n 2) 23–4.

⁷⁹ *Briggs* (n 60) 477 [30] (per McColl JA).

⁸⁰ See Burns articles cited above in n 18.

⁸¹ Tristan W Casey, Xiaowen Hu, Qian Yi Lee and Clarissa Carden, *Stigma Towards Injured or Ill Employees: Research on the Causes and Impact of Stigma in Workplaces, and Approaches to Creating Positive Workplace Cultures that Support Return to Work* (Report for Safe Work Australia, June 2021) <<https://www.safeworkaustralia.gov.au/doc/research-causes-and-impact-stigma-workplaces-and-approaches-creating-positive-workplace>>.

⁸² *Kozarov* (VSC) (n 9) 163 [733]. See also 163 [734]–[739].

⁸³ *Ibid* 163 [733]. Dixon J noted at 163 [733]–[735] that there was no evidence, including evidence led by the OPP, that suggested Kozarov could not have been rotated to another part of the OPP. ‘No explanation was provided [by the OPP] as why a role entirely away from sexual offences ... was not a viable alternative.’: at 166 [750].

⁸⁴ *Kozarov* (VSCA) (n 9) 478 [106].

⁸⁵ *Ibid*.

was aware at that stage she had PTSD.⁸⁶ However, the Court of Appeal ‘looked afresh’ at the evidence.⁸⁷ It relied on her emails in August 2011 where she ‘reacted strongly’⁸⁸ against any suggestion she was not coping and indicated she was passionate about her work. It also relied on her application for and acceptance of promotion. The Court of Appeal found that on the balance of probabilities it could not be satisfied she would have accepted rotation away from the SSOU.

The *Kozarov* case demonstrates the importance of hypothetical factual inferences in determining the counterfactual in causation; that is, what would the plaintiff have done if the breach had not occurred? Kozarov argues that the factual finding by the Court of Appeal is ‘perverse’⁸⁹ and is ‘contrary to common sense’.⁹⁰ She argues that the Court overlooked that her case was not restricted to rotation as the only option to reduce exposure and that there were other means to reduce trauma exposure suggested by the expert witnesses and accepted by Dixon J at the trial.⁹¹ Additionally, Kozarov submits that the counterfactual has to proceed on the basis of what she would have done if she had been aware at the time of the sentinel events that she had PTSD.⁹² Kozarov also argues that reliance on her statements about her passion for her work and her wish to remain at the SSOU is misconceived given unchallenged expert evidence ‘readily explained’ that dedication to the job was in itself a symptom of and consistent with PTSD.⁹³

The factual inferences drawn by the Court of Appeal should be approached with caution. As a matter of principle, the counter-factual inquiry must proceed on the basis that at the time of the hypothetical decision-making about whether to consent to rotation, Kozarov would have been aware she likely had PTSD if the OPP had not been in breach of their duty. Care should also be taken in making inferences about how Kozarov would have acted based on statements made in a wholly different context (that is, an apparent industrial type dispute). Expert evidence in the case suggested that it is not abnormal for people exposed to vicarious trauma and who are not aware they are suffering PTSD to be dedicated to their jobs — this was reason for extra care to be taken by the employer.⁹⁴ It is also not necessarily inconsistent to want to progress in your workplace through promotion and, at the same time, expect your employer to take steps to prevent known risks to your health. There was no evidence at trial that suggested Kozarov could not be rotated or as to why other preventative steps could not be taken to reduce trauma exposure.⁹⁵ The taking of steps such as rotation or altering work allocation would have been consistent with the OPP’s own VT Policy.

⁸⁶ *Kozarov (VSC)* (n 9) 163 [733].

⁸⁷ *Kozarov (VSCA)* (n 9) 479 [110]. See also 478–9 [108]–[109].

⁸⁸ *Ibid* 478 [108].

⁸⁹ Kozarov’s Submissions (n 50) [21].

⁹⁰ *Ibid* [30], [35].

⁹¹ *Ibid* [32], [35].

⁹² Zagi Kozarov, ‘Appellant’s Reply’, Submission in *Kozarov v Victoria*, Case No M36/2021, 27 August 2021, [4] <https://cdn.hcourt.gov.au/assets/cases/06-Melbourne/m36-2021/Kozarov-Vic_Reply.pdf> (‘Kozarov’s Reply’).

⁹³ Kozarov’s Submissions (n 50) [37].

⁹⁴ *Kozarov (VSC)* (n 9) 146 [652], 158 [714]–[715], 159 [718].

⁹⁵ See above n 83 and accompanying text.

The impact of employment contracts on negligence emerges as a significant issue in the High Court. The OPP argues that a duty cannot be imposed that requires an employer to breach the employment contract by rotating an employee without consent and enforcing a safe system of work cannot be inconsistent with contractual obligations.⁹⁶ The OPP submits that Kozarov's employment contract was to be specifically employed in the SSOU, such that any rotation or employment elsewhere in the OPP would be precluded by contract.⁹⁷ This appears to be a substantially new issue not raised at trial.⁹⁸ Following Kozarov's notification of her injury and request for rotation out of the SSOU in February 2012, she was in fact moved to a range of other areas in the OPP. Kozarov submits that the argument is impermissible as there was a concession at trial that there was nothing in the contract that prevented steps being taken to prevent injury.⁹⁹ As the impact of this aspect of the contract of employment has not previously been fully ventilated, it seems incongruous to explore it for the first time in the High Court. There are aspects of the contract of employment that were not substantively discussed in the Court of Appeal. There are quite complex interactions between written employment contracts, enterprise agreements, public service administration legislation, industrial legislation, and OHS and workers' compensation legislation that were not explored and may affect how the employment contract interacts with negligence principles. For example, employment contracts could not generally exclude obligations or benefits contained in industrial, OHS or workers' compensation legislation.¹⁰⁰ In addition, it is not at all clear that if Kozarov was diagnosed with a psychiatric illness affecting her work, she could not have been lawfully moved out of her role as a result of provisions in public service administration legislation, workers' compensation legislation, OHS or industrial legislation.

A further issue raised by Kozarov is whether the effect of the Court of Appeal's findings on causation negates an employer's obligation to enforce a safe system of work and deprives the duty and standard of care of content and effect.¹⁰¹ Kozarov seeks to apply longstanding principles that the employer's duty extends not just to the establishment of a duty to create a safe system of work, but also to the obligation to maintain and enforce the system.¹⁰² Kozarov argues the OPP cannot be excused from fulfilling their duty of care based on any hypothetical wishes of the employee.¹⁰³ The High Court will ultimately resolve these tensions in light of consideration of the employment contract. However, it appears incoherent with the broader OHS and other legislative context for preventative health measures by employers to be overridden by employee wishes.

⁹⁶ Victoria's Submissions (n 63) [47]–[49].

⁹⁷ Ibid [42].

⁹⁸ See above n 83 and accompanying text.

⁹⁹ Kozarov's Reply (n 92) [5].

¹⁰⁰ See, eg, the National Employment Standards in *Fair Work Act 2009* (Cth) pt 2-2 including maximum reasonable work hours (s 62), which may make some of the findings in *Koehler* outdated.

¹⁰¹ Kozarov's Submissions (n 50) [43]–[48].

¹⁰² Ibid [43]–[48] *McLean* (n 17) 155 CLR 306, 313–14.

¹⁰³ Kozarov's Submissions (n 50) [42].

IV Conclusion

The *Kozarov* appeal raises important issues arising from the application of the *Koehler* principles to workplace psychiatric injury cases, particularly those concerning vicarious trauma. These include the interpretation of reasonable foreseeability in cases where an employer has actual knowledge of the risk of psychological injury from vicarious trauma; the role of autonomy and privacy; how to approach factual inferences in counterfactual causation inquiries; and the interaction between employment contracts and an employer's obligation to enforce a safe system of work, and causation. Reflection on the case demonstrates that there are further areas where the *Koehler* principles could be clarified or reassessed either in the *Kozarov* appeal or in future cases. Is there always a need for explicit signs of psychiatric illness from a particular employee, where the employers have clear knowledge of the risk of psychiatric injury to each and every employee due to the traumatic nature of the work?¹⁰⁴ Is it sufficient for an employer to stand by and do nothing to respond to known risks of psychiatric workplace injury by way of a safe system of work, until it is very obvious that injury is manifesting? Has there been too much focus on individual factors that contribute to employee psychiatric injury and too little on well-known systemic workplace factors?¹⁰⁵ To what extent should judges in negligence cases rely on common sense social facts about what employers and employees know and how they behave? *Kozarov* also raises the need to closely examine employment contracts in light of the complex legislative environment that affects employment, and to give proper consideration to coherence between OHS legislative obligations and negligence. At a conceptual level, it is questionable in light of significant contemporary research, to continue to distinguish between physical and psychiatric injury in the workplace. It would be appropriate to recognise a proactive and reactive duty of care to ensure a safe system of work that responds to both kinds of workplace injury (similar to OHS legislative duties) with issues of appropriate reasonable responses in the circumstances dealt with at the breach stage.

¹⁰⁴ There may be an argument that there is sufficient flexibility in the *Koehler* decision to allow this to occur in appropriate circumstances: see above n 12 and accompanying text.

¹⁰⁵ ACL Davies, 'Stress at Work: Individuals or Structures' (2021) *Industrial Law Journal*, wab006 <<https://doi.org/10.1093/indlaw/dwab006>>.

Case Note

Love v Commonwealth: The Section 51(xix) Aliens Power and a Constitutional Concept of Community Membership

Mischa Davenport*

Abstract

Love v Commonwealth represents a significant shift in the High Court of Australia's jurisprudence on s 51(xix) of the *Australian Constitution*. Whereas previous cases have alluded to the existence of theoretical limits to the scope of the s 51(xix) aliens power, the result in *Love v Commonwealth* involves the declaration and enforcement of such a limit in practice, with the majority holding that Aboriginal Australians are beyond the scope of the power. Perhaps more significantly, the majority approach to the aliens power positions the Court to develop a substantive concept of constitutional membership. The minority analyses of s 51(xix) instead adopt a sovereignty framework approach: they proceed on the basis that legislation relying on s 51(xix) can validly apply to any person so long as the criterion attracting its application has a plausible connection to the ordinary understanding of alienage. The minority approach rejects the notion of a constitutional concept of community membership and would instead give the Commonwealth Parliament a broad discretion to determine matters of membership and alienage. It is in the majority's rejection of this approach that a concept of community membership emerges. While Parliament, on the majority view, retains a degree of control over the composition of the constitutional community, its power to exclude persons from that community is subject to significant limitations, including by reference to a concept of territoriality. This case note focuses exclusively on the emergence, in the majority reasons, of a constitutional concept of community membership.

Please cite this case note as:

Mischa Davenport, '*Love v Commonwealth*: The Section 51(xix) Aliens Power and a Constitutional Concept of Community Membership' (2021) 43(4) *Sydney Law Review* 589.



This work is licensed under a Creative Commons Attribution-NonDerivatives 4.0 International Licence (CC BY-ND 4.0).

As an open access journal, unmodified content is free to use with proper attribution. Please email sydneylawreview@sydney.edu.au for permission and/or queries.

© 2021 Sydney Law Review and author. ISSN: 1444-9528

* Graduate Solicitor, NSW Crown Solicitor's Office. Email: mischadavenport@gmail.com. A version of this case note was awarded the 2020 Peter Paterson Prize for best LLB or JD student contribution to the *Sydney Law Review* unit of study. Thank you to my supervisor, Professor Helen Irving, for her valuable input and feedback on earlier drafts.

I Introduction

In *Love v Commonwealth*, the High Court of Australia was asked to determine whether the plaintiff Aboriginal men (Daniel Love and Brendan Thoms) were aliens for the purposes of s 51(xix) of the *Australian Constitution*.¹ Love and Thoms were born in Papua New Guinea and New Zealand respectively, each with an Australian citizen parent, but holding the citizenship of their respective country of birth and never having acquired Australian citizenship. Both had been permanent residents in Australia before cancellation of their visas under s 501(3A) of the *Migration Act 1958* (Cth) and would have been liable to deportation as ‘unlawful non-citizens’ under s 198 of that Act. The power to deport each plaintiff turned exclusively on the Commonwealth’s s 51(xix) power to legislate with respect to ‘[n]aturalization and aliens’.

The novelty of the plaintiffs’ position within the Court’s s 51(xix) jurisprudence was their Aboriginality. The Court was asked to determine whether that status had any constitutional significance for the purposes of the aliens power. Four Justices (Bell, Nettle, Gordon and Edelman JJ)² held that Aboriginal Australians (as defined)³ could not be considered aliens for the purposes of s 51(xix).⁴ However, to limit the significance of *Love* to this narrower proposition would ignore important features of the majority’s overall approach to the aliens power.

The disagreement between the majority and the minority reflects a difference in underlying conceptions of the relationship between individuals and the Commonwealth of Australia that constitutes ‘membership’ of the constitutional community. For the minority, this relationship is essentially formal, regulated by the exercise of a broad legislative discretion to control community membership as citizenship. For the majority, on the other hand, there is a substantive, pre-legislative concept of community membership that is defined by reference to a *particular* community’s assertion of sovereignty over a *particular* territory. This concept of territoriality, featuring to varying degrees in each of the majority judgments, mediates the relationship between the individual and the Commonwealth and may play an important role in the future development of a concept of constitutional membership.

¹ *Love v Commonwealth* (2020) 270 CLR 152, 169–70 [1]–[4] (‘*Love*’).

² Though each Justice in *Love* gave separate reasons, when addressing common features of the reasons of Bell, Nettle, Gordon and Edelman JJ, I will refer to them as reasons of the majority. Similarly, the separate reasons of Kiefel CJ, Gageler and Keane JJ will be collectively referred to as reasons of the minority.

³ While the relevant test of Aboriginality for the purposes of s 51(xix) is an important point of disagreement in *Love*, it is not within the scope of this case note. It has been addressed elsewhere more thoroughly than would be possible here, see Michelle Foster and Kirsty Gover, ‘Determining Membership: Aboriginality and Alienage in the High Court’ (2020) 31(2) *Public Law Review* 105. Practical considerations arising from the need to assess Aboriginality have since come before the Full Court of the Federal Court of Australia: *McHugh v Minister for Immigration, Citizenship, Migrant Services and Multicultural Affairs* (2020) 283 FCR 602.

⁴ The constitutional ‘orthodoxy’ (at 110) of the majority’s interpretive approach to this narrower question has been amply defended by Gerangelos: Peter Gerangelos, ‘Reflections upon Constitutional Interpretation and the “Aliens Power”: *Love v Commonwealth*’ (2021) 95(2) *Australian Law Journal* 109.

This case note explores the concept of constitutional membership that emerges from *Love*. Part II outlines the Court's prior aliens power jurisprudence, setting out the broad limits within which the scope of the power remained to be determined. Part III summarises the Court's reasons in *Love*, including the minority position. Part IV addresses the concept of constitutional membership that emerges from the majority reasons, the relevance of territoriality to that concept, and underlying questions of constitutional legitimacy that may inform the concept's future development.

II Pre-Love Aliens Case Law

Previous High Court decisions concerning the aliens power had, without meaningful exception, affirmed Commonwealth power to exclude various 'outsiders' from the Australian community. In the first half of the 20th century, before the advent of statutory Australian citizenship, the primary power of exclusion was the immigration power in s 51(xxvii) of the *Constitution*.⁵ When the aliens power was discussed, the Court emphasised its breadth, stating that it

must surely, if it includes anything, include the power to determine the conditions under which aliens may be admitted to the country, the conditions under which they may be permitted to remain in the country, and the conditions under which they may be deported from it.⁶

Reliance on the immigration power allowed the Commonwealth to exclude a range of (non-white) British subjects who, as a consequence of Australia's place within the British Empire at the time, would not have been considered aliens for the purposes of s 51(xix).⁷ But even the power of exclusion conferred by s 51(xxvii) was not unlimited in scope. In *Potter v Minahan*, an Australian-born (and thereby British subject) child of an Australian mother and a Chinese father, taken back to China by his father at a young age, was held, when returning to Australia as an adult, not to be an immigrant, but rather 'a member of the [Australian] community'.⁸ Thus where immigrant (rather than alien) status was relied on as the basis for exclusion, Commonwealth power was limited by reference to a concept of community membership.

As Australian citizenship replaced British subject status,⁹ the Commonwealth relied increasingly on s 51(xix) rather than s 51(xxvii). A series of three High Court cases on the aliens power addressed the changing significance of British subject

⁵ See *Potter v Minahan* (1908) 7 CLR 277; *Attorney-General (Cth) v Ah Sheung* (1906) 4 CLR 949.

⁶ *Robtelves v Brennan* (1906) 4 CLR 395, 404 (Griffith CJ).

⁷ See Sangeetha Pillai, 'Non-Immigrants, Non-Aliens and the People of the Commonwealth: Australian Constitutional Citizenship Revisited' (2013) 39(2) *Monash University Law Review* 568, 579–83; John Gava, 'Losing Our Birthright: *Singh v Cth*' (2016) 37(2) *Adelaide Law Review* 369, 378. See also *R v Macfarlane; Ex parte O'Flanagan* (1923) 32 CLR 518, 556 (Isaacs J): 'the power as to "aliens" leaves a huge gap, sufficient in itself to paralyse the Commonwealth unless "immigration" covers it'.

⁸ *Potter v Minahan* (n 5) 289 (Griffith CJ). See also at 299 (Barton J): 'No one describes a man returning home to his own country as an immigrant. ... Immigration has various but kindred meanings. They all imply that the country which the immigrant seeks to enter is not his home, by any criterion, natural or artificial.'

⁹ See especially the *Nationality and Citizenship Act 1948* (Cth), and changes made to it by the *Australian Citizenship Amendment Act 1986* (Cth).

status across Australia's gradual trajectory towards independent sovereignty.¹⁰ All concerned plaintiffs who, having entered Australia as British subjects and not having thereafter acquired Australian citizenship, argued that by virtue of their British subject status at the moment each entered Australia (between 1966 and 1974) they could not be considered aliens. *Re Patterson; Ex parte Taylor* stands out, with the High Court holding the plaintiff to be outside the scope of the aliens power.¹¹ However, it was effectively overruled in *Shaw v Minister for Immigration and Multicultural Affairs*, in which a majority held that, from the commencement of the *Australian Citizenship Act 1948* (Cth), a British subject born outside Australia to non-citizen parents entered Australia as an alien and would remain so unless and until naturalised.¹²

More recent aliens cases, not concerned with British subject status, have without exception affirmed Commonwealth power to exclude non-citizen plaintiffs, while nonetheless acknowledging theoretical limits on that power. In *Pochi v MacPhee*, two relevant points were made concerning the aliens power. First, while the meaning of 'aliens' must be ascertained by reference to Australian rather than foreign law,

[c]learly the Parliament cannot, simply by giving its own definition of 'alien', expand the power under s 51(xix) to include persons who could not possibly answer the description of 'aliens' in the ordinary understanding of the word.¹³

Second, unlike immigrant status, alien status is not subject to a principle of absorption: a person does not, merely by reason of long residence in Australia, cease to be an alien.¹⁴ The only way for an alien to be relieved of that status is by naturalisation,¹⁵ a process exclusively regulated by Commonwealth legislation.¹⁶ More recent High Court rulings have confirmed this point.¹⁷

Following legislative abrogation of the common law *ius soli* principle,¹⁸ a wide range of non-citizens have been found to fall within the scope of the aliens

¹⁰ See *Nolan v Minister for Immigration and Ethnic Affairs* (1988) 165 CLR 178 ('Nolan'); *Re Patterson; Ex parte Taylor* (2001) 207 CLR 391 ('Re Patterson'); *Shaw v Minister for Immigration and Multicultural Affairs* (2003) 218 CLR 28 ('Shaw').

¹¹ *Re Patterson* (n 10).

¹² *Shaw* (n 10) [32] (Gleeson CJ, Gummow and Hayne JJ). The majority reasoning in *Shaw* was recently endorsed and expanded upon by a majority of the High Court in *Chetcuti v Commonwealth* (2021) 95 ALJR 704 ('Chetcuti'). In *Chetcuti*, the majority held that a British subject, having arrived in Australia in 1948 (before the commencement of the *Australian Citizenship Act 1948* (Cth) on 26 January 1949) and not thereafter having acquired Australian citizenship was susceptible to treatment as an alien.

¹³ *Pochi v MacPhee* (1982) 151 CLR 101, 109 (Gibbs CJ) ('Pochi'). See also *Re Patterson* (n 10) 400 [7] (Gleeson CJ): 'Whilst fully accepting that Parliament cannot, by some artificial process of definition, ascribe the status of alienage to whomsoever it pleases ...'.

¹⁴ See *Pochi* (n 13) 113 (Murphy J).

¹⁵ Registration of citizenship by descent may, for constitutional purposes, be considered a form of naturalisation.

¹⁶ While naturalisation in the *Australian Citizenship Act 2007* (Cth) is referred to as 'citizenship by conferral' (pt 2 div 2 sub-div B), I will, for convenience, continue to use the generic term 'naturalisation'.

¹⁷ See, eg, *Re Minister for Immigration and Multicultural Affairs; Ex parte Te* (2002) 212 CLR 162, 172 [26] (Gleeson CJ) ('Ex parte Te').

¹⁸ The *Australian Citizenship Amendment Act 1986* (Cth) (n 9) restricted automatic citizenship to children born in Australia to at least one citizen or permanent resident parent.

power, including children born in Australia (to non-citizen, non-resident parents) but holding a foreign citizenship,¹⁹ and even children born in Australia holding no citizenship at all.²⁰ Despite regular reference to the existence of theoretical limits of the kind alluded to in *Pochi*, prior to *Love* the High Court had never enforced such limits in practice. As a result, the case law does not reveal any particular feature capable of taking a person beyond the scope of the power. Absorption into the Australian community, birth within Australia, or the absence of foreign allegiance had not sufficed.²¹ In the particular circumstances of Papua New Guinea's independence from Australia, even a person born an Australian citizen, to Australian citizen parents, within the then territory of Australia, was not safe from the operation of the aliens power following independence.²²

However, while no feature had emerged in the High Court jurisprudence as definitive of non-alienage, there was similarly no obvious candidate for a definitive criterion of alienage. The only feature common to all of the plaintiffs held to have been validly excluded under the aliens power was their lack of Australian citizenship, but if the *Pochi* limits are to have any substance, then non-citizenship alone cannot be determinative: citizenship has no constitutional status and is subject to legislative modification. Thus a challenge for the Court in *Love*, regardless of the position taken with respect to Aboriginality, was to articulate satisfactorily the *Pochi* limits in a way that previous decisions had failed to do.

III The High Court in *Love*

Each of the separate judgments in *Love* accepted the existence of limits on the aliens power of the kind alluded to in *Pochi*.²³ Moreover, all agreed that the aliens power is not subject to a doctrine of absorption of the kind developed in the context of the s 51(xxvii) immigration power,²⁴ and that alien status can only be lost by naturalisation.²⁵ These propositions from *Pochi* set the scene for the analysis of s 51(xix): on the one hand Parliament's power must be limited by reference to some substantive concept of alienage, yet at the same time the concept has a formal element at least to the extent that an alien may only be relieved of the status by the (formal) process of naturalisation, as provided for by Commonwealth legislation.

¹⁹ *Singh v Commonwealth* (2004) 222 CLR 322 ('*Singh*').

²⁰ *Koroitamana v Commonwealth* (2006) 227 CLR 31 ('*Koroitamana*').

²¹ Note that a plausible argument has been made that the Court in *Singh* (n 19) broke with established precedent in declaring that birth in Australia would not take a person beyond the scope of the aliens power: Gava (n 7); Anthony Gray, 'The Meaning of an "Alien" in the Constitutional Universe' (2013) 20(2) *Australian Journal of Administrative Law* 89. Despite the re-emergence of a concept of community membership in *Love*, there is no suggestion that *Singh* is likely to be reconsidered. However, as the plaintiffs in *Love* were born overseas, the question was not squarely raised.

²² See *Re Minister for Immigration and Multicultural and Indigenous Affairs; Ex parte Ame* (2005) 222 CLR 439 ('*Ex parte Ame*'), but note that the reasoning in that case was substantially concerned with the interaction of the aliens power with the territories power in s 122 of the *Constitution* in its application to external territories.

²³ *Love* (n 1) 171–2 [7] (Kiefel CJ), 183 [50]–[51] (Bell J), 194–5 [87] (Gageler J), 218 [168] (Keane J), 236–7 [236], 238–9 [242] (Nettle J), 266 [310] (Gordon J), 288 [394] (Edelman J).

²⁴ See, eg, *ibid* 175 [19] (Kiefel CJ), 246 [257] (Nettle J), 264 [304] (Gordon J).

²⁵ See *ibid* 174 [17] (Kiefel CJ), 247–8 [261] (Nettle J), 264 [304] (Gordon J). But note at 299–301 [416]–[421] Edelman J argues that absorption may be a 'relevant factor' (at 319 [464]) in the determination of non-alien status, suggesting a willingness to reconsider *Pochi* (n 13) on this point.

The disagreement in *Love* concerned not only the application of the *Pochi* limits to the plaintiffs, but also the far more significant question of the nature of those limits generally. On the latter question, the minority approach would attribute significant discretionary power to Parliament to determine the composition of the constitutional community, while the majority approach more significantly restricts that discretion. The difference between the two approaches is set out below.

A *Framing the Pochi Limits on Section 51(xix) Legislative Power*

For the majority in *Love*, the ‘ordinary understanding’²⁶ of alienage limits the persons, or categories of persons, to whom legislation supported by s 51(xix) can validly apply. Thus, the question in applying the *Pochi* limits is whether a particular individual (to whom a law supported by s 51(xix) purportedly applies) is in fact an alien, or capable of answering the description of ‘alien’ in the ordinary understanding of the word.²⁷ That inquiry is primarily concerned with elements of an individual’s status capable of taking him or her outside the ordinary understanding of alienage, and thus beyond the scope of the power.

For the minority in *Love*, on the other hand, the ordinary understanding of alienage only limits the criteria by reference to which Parliament may legislatively attribute the status. Thus, the question of validity instead turns on features of the law itself, rather than features of the persons to whom it applies. A law would validly *determine* a certain class of persons to be aliens if the criterion or criteria for the attribution of that status bore a sufficient connection to the ordinary understanding of alienage. Essentially, for the minority there is no pre-legislative fact of a person’s alienage, merely valid and invalid criteria by reference to which Parliament may attribute that status. The power to regulate community membership is only limited by the requirement that the law identifying non-members must fasten on some feature bearing a sufficient connection to the ordinary understanding of alienage. Kiefel CJ framed the question as ‘whether it is open to the Commonwealth Parliament to treat persons having the characteristics of the plaintiffs as non-citizens [aliens] for the purposes of the *Migration Act*’.²⁸ Accordingly, for the minority, ‘the status of alien is not defined by pointing to what is said to take a person outside the reach of Parliament’s prescription, rather it depends upon what it is that gives the person that status’²⁹ — the latter criteria being those set out by Parliament, in negative form, in the *Australian Citizenship Act 2007* (Cth).

On this minority view, once the validity of a chosen criterion of ‘alienage’ has been established, it is irrelevant that an individual so identified may in some other sense be considered a member of the community.³⁰ The *Pochi* limits would simply mean that ‘there are “available characteristics for the Parliament to choose

²⁶ *Love* (n 1) 183 [50]–[51] (Bell J), 236–7 [237] (Nettle J), 300 [419] (Edelman J).

²⁷ See, most succinctly, *ibid* 261 [293] (Gordon J).

²⁸ *Ibid* 170 [4] (Kiefel CJ). Note that for the minority Justices, for whom valid Commonwealth legislation is determinative of alien status, citizenship and alienage are taken to be mutually exclusive antonyms: see, eg, *ibid* 170–1 [5] (Kiefel CJ), 219–20 [172] (Keane J).

²⁹ *Ibid* 174 [15] (Kiefel CJ).

³⁰ *Ibid* 220 [172] (Keane J): ‘the fact that a person who is not a citizen of Australia also has some other characteristic (such as having been born to an Australian parent, or having deep personal ties or a strong emotional attachment to Australia) cannot alter that status created by law’.

and some unavailable characteristics”³¹. It would, thus, be open to Parliament to pass a law attributing alien status to any class of individuals so long as the characteristic by which that class was identified was related to the ordinary concept of alienage. Presumably, this would enable Parliament to treat as aliens persons possessing any one or a combination of characteristics plausibly connected to the concept of alienage, including at the very least birth outside the Australian territory, birth to one or more non-citizen parents, or foreign allegiance.

Gageler J, defending this approach, describes alienage as a ‘recognized topic of juristic classification’,³² meaning that the power to legislate with respect to aliens necessarily includes ‘a source of legislative authority to modify or replace the pre-existing law on that topic’.³³ The basis for this characterisation is the proposition that the concept of alienage has never had an ‘established and immutable legal meaning’,³⁴ but rather has always been determined by the application of positive law on the subject from time to time.

This analysis, with respect, appears to beg the question of whether past determinations of alien status have, in fact, depended on positive law or, instead, on an underlying substantive concept of alienage or community membership. Where positive law overlaps with an underlying constitutional concept, the difference will be difficult to discern. Much is made by the minority in *Love* of statements in past aliens cases to the effect that s 51(xix) includes a power to ‘determine’ who is an alien: Kiefel CJ describes this as a ‘power to choose the criteria for alienage’;³⁵ Keane J describes s 51(xix) as empowering the Commonwealth ‘to “create and define the concept of Australian citizenship”, to select or adopt the criteria for citizenship or alienage, and to attribute to any person who lacks the qualifications for citizenship “the status of alien”’;³⁶ Gageler J describes s 51(xix) as encompassing a power ‘to determine who is and who is not to have the legal status of alienage’.³⁷

The difficulty with these propositions is that they do not specify exactly how Parliament may determine who will be treated as an alien or a community member. Authorities relied upon by the minority may mean nothing more than that the Parliament has absolute control over the process (naturalisation) by which aliens become members of the community. Gleeson CJ and Heydon J in *Koroitamana v Commonwealth* stated that ‘[t]he power conferred by s 51(xix) is a wide power, under which the Parliament has the capacity to decide who will be *admitted* to formal membership of the Australian community, which now means citizenship’.³⁸ Gleeson CJ in *Re Minister for Immigration and Multicultural Affairs; Ex parte Te* stated that ‘the power to make laws with respect to aliens has been understood as a

³¹ Ibid 186 [60] (Bell J), summarising the Commonwealth’s submission on this point.

³² Ibid 194 [86] (Gageler J), quoting *Attorney-General (Vic) v Commonwealth* (1962) 107 CLR 529, 578 and *Commonwealth v Australian Capital Territory* (2013) 250 CLR 441, 455 [14].

³³ *Love* (n 1) 194 [86] (Gageler J).

³⁴ Ibid, quoting *Koroitamana* (n 20) 37 [9].

³⁵ *Love* (n 1) 170 [5] (Kiefel CJ), citing *Koroitamana* (n 20) 38 [11].

³⁶ *Love* (n 1) 217 [166] (Keane J), citing *Koroitamana* (n 20) 37 [9], 46 [48], 46 [50], 49 [62] and *Shaw* (n 10) 35 [2].

³⁷ *Love* (n 1) 193 [84] (Gageler J), citing *Ex parte Te* (n 17) 170–2 [21]–[26] (Gleeson CJ), 219–20 [209]–[210] (Hayne J) and *Shaw* (n 10) 35 [2], 87 [190].

³⁸ *Koroitamana* (n 20) 38 [11] (emphasis added).

wide power, equipping the Parliament with the capacity to decide, on behalf of the Australian community, who will be *admitted* to formal membership of that community'.³⁹ The more limited proposition, that Parliament has power to regulate community membership via its control over the process of naturalisation, does not deny the existence of an underlying concept of community membership independent of positive law on the subject.

For the minority in *Love*, however, the broader power to regulate community membership is seen as a key element of State sovereignty. For Gageler J, '[m]embership of or exclusion from the political community of the Commonwealth of Australia is a topic of vital national importance'.⁴⁰ For Kiefel CJ, it is 'a serious matter' to deny to Parliament a broader power to regulate community membership that 'is fundamental to the structure of the *Constitution* and the governance of Australia'.⁴¹ The language used is reminiscent of an implied nationhood power.⁴² However, there are counter-examples to the proposition that a broad legislative discretion to determine community membership is an essential element of national sovereignty.⁴³

Within this 'sovereignty framework', the power of exclusion would only be constrained by the minimal requirement that the exclusion of a given class of persons be linked to a characteristic plausibly connected to alienage. Previous aliens cases would thus establish the validity of relevant characteristics as criteria for the legislative denial of membership. On the alternative (*Love* majority) view, Commonwealth power to control the composition of the constitutional community is implicitly limited to its regulation of the process of naturalisation. The power of exclusion is constrained by a constitutional concept of community membership, with members of that constitutional category not capable of answering the description of aliens. Previous aliens cases could only affirm that each particular plaintiff, validly subject to the aliens power, was not, in fact, a member of the community in the constitutional sense. In the next section, I consider how these different conceptions of the limits on the s 51(xix) aliens power were applied to the plaintiffs in *Love*.

B Applying the Pochi Limits to the Plaintiffs in *Love*

The majority's analysis in *Love* focuses on ways in which the plaintiffs belong to the Australian community, these factors being capable of taking them outside the ordinary understanding of aliens.⁴⁴ The minority, while raising the plaintiffs' lack of Australian citizenship, can be most coherently interpreted as focusing on ways in

³⁹ *Ex parte Te* (n 17) 171 [24] (emphasis added).

⁴⁰ *Love* (n 1) 209 [130].

⁴¹ *Ibid* 173 [14]. See also at 217–18 [167] (Keane J): 'What was clear at Federation was that it was an attribute of the sovereignty of an independent State to decide who were aliens and whether they should or should not become members of the community.'

⁴² See, eg, *Ruddock v Vadarlis* (2001) 110 FCR 491.

⁴³ See, eg, the United States, where a constitutionally enshrined concept of citizenship (in the *United States Constitution* amend XIV) strictly limits Congress' power: *Afroyim v Rusk*, 387 US 253 (1967); Patrick Weil, 'Can a Citizen Be Sovereign?' (2017) 8(1) *Humanity: An International Journal of Human Rights, Humanitarianism, and Development* 1.

⁴⁴ See, eg, *Love* (n 1) 263 [302] (Gordon J): 'The word "alien" ... describes a person's "lack of relationship with a country"' (emphasis in original), citing *Nolan* (n 10) 183, quoted in *Singh* (n 19) 400 [205].

which the plaintiffs are linked to foreign places to bring them within the ordinary understanding of alienage.⁴⁵

The minority judgments are more uniform and so may be treated together. Their starting point in the search for valid criteria of alienage is the meaning of the term ‘alien’ at Federation. According to Gageler J,

it must now be taken as settled that the Parliament is entitled at least to choose between the principal options recognised as having vied for acceptance as indicia of nationality in the second half of the 19th century, being place of birth (*jus soli*) or the nationality of one or more parents (*jus sanguinis*), or to choose some combination of the two.⁴⁶

The result is that at least birth outside the territory (the obverse of *ius soli*) and birth to a non-citizen parent (the obverse of *ius sanguinis*) must be valid criteria of alienage. By conferring automatic citizenship based on the conjunction of *ius sanguinis* and *ius soli* rules, Parliament has left both criteria as determinants of alienage, so that birth outside of the territory or to a non-citizen parent would be a valid basis for the Commonwealth to treat each plaintiff as an alien.

The minority also places significant weight on the question of foreign allegiance, relying on the statement in *Singh v Commonwealth* that ‘a central characteristic of the status of “alien” is, and always has been, owing obligations to a sovereign power other than the sovereign power in question’.⁴⁷ Care is taken, however, to emphasise that it is not a necessary condition of alienage (in recognition of the Court’s earlier decision in *Koroitamana*),⁴⁸ nor a sufficient condition (as it is within Parliament’s power to provide, as it has done, for dual citizenship).⁴⁹ Nonetheless, the minority suggests that foreign allegiance would be a valid criterion for the legislative attribution of alienage, even though it is not independently determinative.⁵⁰ Importantly, the allegiance in question is merely formal, describing the relation of a person to the State of which he or she is a citizen or subject.⁵¹

While accepting that the content of a concept of allegiance is unclear,⁵² the minority at times appears to rely on the *absence* of formal allegiance (citizenship) as a determinative criterion of alienage.⁵³ However, this should be interpreted in light of the *Pochi* limits, which would be non-existent if the legislative capacity with respect to citizenship were not correspondingly limited. Thus, while the absence of formal allegiance may, presuming the validity of citizenship laws, conclusively indicate alien status, it is not readily characterised as a valid criterion for the legislative attribution of that status.

⁴⁵ *Love* (n 1) 175 [18] (Kiefel CJ): ‘as a matter of etymology, “alien” means belonging to another place’. See also at 197 [93] (Gageler J).

⁴⁶ *Ibid* 200 [100] (Gageler J). See also at 171–2 [6]–[7] (Kiefel CJ), 217–18 [167] (Keane J).

⁴⁷ *Ibid* 218 [169] (Keane J), quoting *Singh* (n 19) 383 [154] (Gummow, Hayne and Heydon JJ).

⁴⁸ *Love* (n 1) 195–6 [89] (Gageler J), 219 [170] (Keane J).

⁴⁹ *Ibid* 219 [171] (Keane J).

⁵⁰ *Ibid* 174 [16] (Kiefel CJ); 195–6 [89] (Gageler J); 219 [170] (Keane J).

⁵¹ *Ibid* 220 [174] (Keane J).

⁵² See, eg, *ibid* 173 [13] (Kiefel CJ), 203–4 [109] (Gageler J).

⁵³ See, eg, *ibid* 178–9 [32]–[33] (Kiefel CJ).

The minority's understanding of s 51(xix) and its limits places the plaintiffs squarely within the scope of the power. Australian citizenship law reflects a valid legislative choice to allow birth outside the territory (among other criteria) to be a determinant of alien status, and that alone is sufficient to bring the plaintiffs within the scope of the power. Foreign allegiance represents a further criterion that Parliament might validly have seized upon to justify its attribution of alien status to the plaintiffs.

The majority judgments in *Love* are more varied in their application of s 51(xix) and its limits to the plaintiffs. All at least implicitly reject the valid-criteria-of-alienage approach to the extent that they do not treat previous aliens cases as establishing criteria capable of bringing a person *within* the scope of the power. Instead, each of the majority reasons approaches the *Pochi* limits by asking whether there is a relevant *connection* to Australia — that is, a relevant indicator of community membership — by reason of which the plaintiffs could not possibly answer the description of aliens.

Bell J starts by observing that, at Federation, Aboriginal Australians were not aliens, though it is an open question whether their non-alienage was by reason of their birth within the territory (and consequent British subject status) or by reason of the more fundamental circumstance of 'the unique connection that Aboriginal Australians have to the land and waters of Australia'.⁵⁴ Bell J notes, however, that the meaning of 'alien', regardless of its meaning at Federation, may change over time in response to 'changes in the national and international context' — noting, in particular, that the apparent power of the Commonwealth to redefine alien status, as emerging from *Shaw* and *Re Minister for Immigration and Multicultural and Indigenous Affairs; Ex parte Ame*, must be understood in this light.⁵⁵

In this connection, Bell J draws attention to *Mabo v Queensland (No 2)*⁵⁶ as involving recognition by the common law of important circumstances — the 'antecedent rights and interest in the land and waters of Australia possessed by the indigenous inhabitants sourced in traditional law and customs'⁵⁷ — that bear significantly on the ordinary understanding of alienage and community membership. It is this underlying recognition of a spiritual connection to country that makes it impossible for Bell J, relying most directly on the *Pochi* formulation, to find 'that an Aboriginal Australian can be described as an alien within the ordinary meaning of that word'.⁵⁸

Nettle J approaches the *Pochi* question via the more traditional concepts of allegiance and protection: the essential meaning of alienage concerns an *absence* of permanent allegiance to the sovereign.⁵⁹ Allegiance in this context, however, is more than merely formal. While the precise content of permanent allegiance defies satisfactory definition,⁶⁰ any person incapable of being classified as an alien must be

⁵⁴ Ibid 183 [52] (Bell J).

⁵⁵ Ibid 189 [69] (Bell J).

⁵⁶ *Mabo v Queensland (No 2)* (1992) 175 CLR 1 (*'Mabo (No 2)'*).

⁵⁷ *Love* (n 1) 189 [70] (Bell J).

⁵⁸ Ibid 189 [71] (Bell J).

⁵⁹ Ibid 240–3 [246]–[248] (Nettle J).

⁶⁰ Ibid 244 [251], 246–7 [257]–[259] (Nettle J).

in possession ‘of characteristics which so connect him or her to the sovereign as necessarily to give rise to reciprocal obligations of protection and allegiance’.⁶¹ The meaning of ‘permanent protection’, while similarly problematic, must be informed by an understanding of the liability of those not entitled to such protection to removal from Australia as aliens.⁶² Common law recognition of rights and interests founded on the existence of Aboriginal societies, with their own traditional laws and customs, imparts an obligation on the sovereign not to ‘tear the organic whole of the society asunder’⁶³ by subjecting its members to a liability to deportation — an obligation amounting to a form of ‘permanent protection’ owed by the sovereign.⁶⁴ Permanent allegiance, by reason of which Aboriginal Australians cannot be said to be aliens, is simply a statement of the counterpart of this permanent protection.⁶⁵

Gordon J’s approach to the *Pochi* question more explicitly considers the relationship between sovereignty and territory. An alien, in the most basic sense, is a person who is not a member of ‘the people of Australia’.⁶⁶ The meaning of that phrase must be understood by reference to the political community from which the Commonwealth derives its popular sovereignty.⁶⁷ That sovereignty is asserted over a particular territory,⁶⁸ to which the common law recognises Aboriginal Australians as having a unique connection as its ‘first peoples’.⁶⁹ The people of Australia, in whose name that sovereignty is asserted, cannot be thought to exclude the first peoples whose connection to the territory has not been severed.⁷⁰ The ordinary understanding of alienage, for Gordon J, is thus directly concerned with questions of territoriality, sovereignty and legitimacy.

Edelman J takes a similar approach to Gordon J, highlighting the significance of territoriality to the concepts of sovereignty and political community that inform the ordinary understanding of alienage.⁷¹ While Edelman J makes most frequent reference to spiritual or ‘metaphysical ties’⁷² to territory (as identifying ‘belongers’⁷³ to the political community), his Honour also attempts to ground these notions by identifying a basic norm recognised by both statute and the common law. His Honour argues that it is in recognition of this metaphysical attachment to country that the combination of *ius soli* and *ius sanguinis* has never been doubted as giving rise to membership of a political community.⁷⁴

While each member of the majority in *Love* develops the concept differently, all at least implicitly articulate the *Pochi* limits by reference to a concept of

⁶¹ Ibid 247 [260] (Nettle J).

⁶² Ibid 254 [273] (Nettle J).

⁶³ Ibid 254 [272] (Nettle J).

⁶⁴ Ibid 254 [273] (Nettle J).

⁶⁵ Ibid 258 [279] (Nettle J).

⁶⁶ Ibid 277 [351] (Gordon J).

⁶⁷ Ibid 277–8 [351]–[352] (Gordon J).

⁶⁸ Ibid 276–7 [347]–[350] (Gordon J).

⁶⁹ Ibid 260 [289], 274 [340] (Gordon J).

⁷⁰ Ibid 276–7 [349] (Gordon J).

⁷¹ Ibid 308 [438] (Edelman J).

⁷² Ibid 289 [396], 308–9 [438]–[439], 320 [466] (Edelman J).

⁷³ Ibid 288 [394], 308 [437]–[438] (Edelman J).

⁷⁴ Ibid 308–9 [439], 311–12 [443]–[445] (Edelman J).

community membership: a form of connection to Australia by reason of which a person cannot be considered an alien. The significance of the emergence of a constitutional concept of community membership, and the role of territoriality within that concept, will be discussed in the following section.

IV A New Constitutional Concept of Community Membership

The existence of a constitutional concept of community membership — sometimes described as ‘constitutional citizenship’ — has long been discussed in the Australian context.⁷⁵ Prior to *Love*, the scope of such a concept had been considered almost entirely at large,⁷⁶ in no small part due to ‘the High Court’s reticence to give the aliens power an autonomous meaning’.⁷⁷ Previous discussions have focused on the aliens and immigration powers as well as constitutional references to ‘the People of the Commonwealth’.⁷⁸ With respect to the latter, much has been made of sparse judicial pronouncements to the effect that the Parliament’s discretion to regulate community membership (as citizenship) is subject to the requirement that ‘it does not exclude from citizenship *those persons who are undoubtedly among “the people of the Commonwealth”*’.⁷⁹

The reasoning in *Love* provides a starting point for answering two key questions about a constitutional concept of community membership: (A) who is included within the category and on what basis?, and (B) what consequences flow from inclusion within the category?

A Who is Included?

The clear ratio decidendi of *Love* is that at least those persons satisfying the tripartite test of Aboriginality (developed in *Commonwealth v Tasmania*⁸⁰ and applied by the Court in *Mabo (No 2)*⁸¹) are constitutional community members and thus non-

⁷⁵ See, eg, David A Wishart, ‘Allegiance and Citizenship as Concepts in Constitutional Law’ (1986) 15(4) *Melbourne University Law Review* 662; Genevieve Ebbeck, ‘A Constitutional Concept of Australian Citizenship’ (2004) 25(2) *Adelaide Law Review* 137; Michelle Foster, ‘Membership in the Australian Community: *Singh v The Commonwealth* and Its Consequences for Australian Citizenship Law’ (2006) 34(1) *Federal Law Review* 161; Christopher Tran, ‘New Perspectives on Australian Constitutional Citizenship and Constitutional Identity’ (2012) 33 *Adelaide Law Review* 199; Pillai (n 7); Elisa Arcioni, ‘“The People” in the *Australian Constitution*: More than a Vague and Emotionally Powerful Abstraction’ (PhD Thesis, University of Sydney, 2015) especially 344–9; Elisa Arcioni, ‘The Core of the Australian Constitutional People — “The People” as “The Electors”’ (2016) 39(1) *University of New South Wales (UNSW) Law Journal* 421. For a comparative, international perspective, see Jo Shaw, *The People in Question: Citizens and Constitutions in Uncertain Times* (Bristol University Press, 2020).

⁷⁶ Pillai (n 7) 609; Ebbeck (n 75) 164; Foster (n 75) 182–3.

⁷⁷ Michelle Foster, ‘“An ‘Alien’ by the Barest of Threads” — The Legality of the Deportation of Long-Term Residents from Australia’ (2009) 33(2) *Melbourne University Law Review* 483, 540.

⁷⁸ For the most comprehensive overview of potential constitutional sources of a membership concept, see Pillai (n 7).

⁷⁹ *Hwang v Commonwealth* (2005) 80 ALJR 125, 130 [18] (McHugh J) (emphasis added).

⁸⁰ *Commonwealth v Tasmania* (1983) 158 CLR 1, 272–3 (Deane J).

⁸¹ *Mabo (No 2)* (n 56).

aliens.⁸² Whether this constitutional status may extend to other persons or groups on the same or analogous bases remains unclear.

Bell and Gordon JJ each allude to the ‘*sui generis*’ nature of Aboriginality,⁸³ but this alone may not be conclusive. It may be significant for the future development of constitutional community membership that the concept of allegiance — the supposed historical basis of community membership — is recognised as devoid of meaningful content.⁸⁴ Even Nettle J, using the term to refer to something more than the merely formal relationship of citizenship, makes clear that it only describes an entitlement to permanent protection⁸⁵, lending support to the proposition that ‘[a]llegiance merely describes, rather than defines political obligation [of community membership] ... It does not answer the questions of when is a person a member’.⁸⁶

In the absence of any meaningful concept of allegiance, a territorial principle emerges as a potential basis of constitutional community membership. Bell J’s ordinary-understanding approach provides little assistance in characterising the kind of connection to territory that might suffice. Nettle J’s approach highlights a form of connection entitling persons to the permanent protection of the Australian Crown — only a slim starting point for future arguments seeking to expand constitutional membership. Gordon and Edelman JJ’s approaches, framed in terms of legitimacy conditions for the exercise of sovereignty over a particular territory, provide an alternative, but similarly vague, basis on which constitutional membership may be extended.

A starting point might be to locate the territorial principles arising from Nettle, Gordon and Edelman JJ’s judgments within a rights framework. For Nettle J, this is relatively straightforward: there are rights, arising from certain connections to territory, that entitle persons or groups to a form of permanent protection by the State (immunity from removal or exclusion from the territory) in a way that limits legislative power.⁸⁷ For Gordon and Edelman JJ, locating their reasons within a rights framework requires some unpacking of the relationship between sovereign legitimacy and territory, starting with the observation that the assertion of sovereignty over a territory involves a claim to exclusivity (a right to exclusive jurisdiction within that territory, and a right to exclude outsiders from it).⁸⁸

⁸² The minority is critical of this test for non-alienage: *Love* (n 1) 176–7 [23]–[26] (Kiefel CJ), 211–12 [137]–[138] (Gageler J), 225–6 [196]–[198] (Keane J). The majority is divided as to the necessity of the test: compare 253–4 [272] (Nettle J), 282 [367] (Gordon J), with 192 [80] (Bell J), 317 [458] (Edelman J). The question is more thoroughly addressed by Foster and Gover (n 3).

⁸³ *Love* (n 1) 190 [74] (Bell J), 272 [333] (Gordon J).

⁸⁴ This is more emphatically confirmed in *Chetcuti* (n 12), in which a majority of the Court supported the proposition that, upon the introduction of Australian citizenship, the concept of allegiance ‘was altogether swept away’: 712 [21] (Kiefel CJ, Gageler, Keane and Gleeson JJ), 721 [64] (Edelman J), both judgments quoting Clive Parry, *Nationality and Citizenship Laws of The Commonwealth and of The Republic of Ireland* (Steven & Sons, 1957) 92.

⁸⁵ See *Love* (n 1) 247 [260] (Nettle J).

⁸⁶ Wishart (n 75) 706.

⁸⁷ *Love* (n 1) 244 [251]–[252] (Nettle J).

⁸⁸ *Ibid* 276–9 [347]–[357] (Gordon J), 308–12 [438]–[446] (Edelman J).

Addressing the legitimacy of such exclusivity, constitutional theorists have argued that ‘constitutional legitimacy is not self-standing’⁸⁹ — that is, the popular (democratic) sovereignty of an internally self-defining political community may be limited by rights-based considerations. One proposed condition for the legitimation of exclusion is that it not violate essential rights of non-members. Some have argued that this may be satisfied by the existence of international regimes for the avoidance of statelessness and the provision of assistance to refugees so that every person has access to a territory ‘where, at the very least, his or her rights are not violated in a serious way’.⁹⁰ However, this minimalist approach has increasingly been superseded by international human rights regimes recognising particular rights, the enjoyment of which requires access to a particular territory, as imposing limits on State powers of exclusion.⁹¹

These developments reflect a recognition that legitimate State sovereignty depends on both internal legitimacy, derived from popular constituent power, and external legitimacy based on non-violation of essential rights.⁹² While a domestic court is not in a position to challenge sovereign legitimacy,⁹³ it may nonetheless interpret the *Constitution* by reference to considerations of legitimacy. Gordon and Edelman JJ’s references to popular sovereignty and the conditions of its exercise may be read in this light: the kind of connection to territory on which constitutional community membership is founded includes any connection the denial of which would involve such a fundamental violation of rights that it would undermine the sovereignty asserted in the *Constitution*. While this bar may be rather high, it is not impossible that the category thus defined might extend beyond Aboriginal Australians.

However, the more recent decision in *Chetcuti v Commonwealth* casts significant doubt on the future of the concept of constitutional community membership developed in the separate majority reasons in *Love*. In *Chetcuti*, the minority Justices in *Love* were joined by the newest member of the Court, Gleeson J, to form a new majority in support of the proposition that ‘the aliens power encompasses both power to determine who is and who is not to have the legal status of an alien and power to attach consequences to that status’.⁹⁴ While briefly acknowledging the existence of ‘an exception in respect of a person who is an Aboriginal Australian according to the tripartite test in *Mabo v Queensland (No 2)*’,⁹⁵

⁸⁹ Mattias Kumm, ‘The Cosmopolitan Turn in Constitutionalism: An Integrated Conception of Public Law’ (2013) 20(2) *Indiana Journal of Global Legal Studies* 605, 612.

⁹⁰ *Ibid* 619.

⁹¹ See, eg, the European Court of Human Rights’ jurisprudence on the right to private and family life, as discussed by Daniel Thym, ‘Respect for Private and Family Life under Article 8 ECHR in Immigration Cases: A Human Right to Regularize Illegal Stay?’ (2008) 57(1) *International and Comparative Law Quarterly* 87; Georgios Milios, ‘The Immigrants’ and Refugees’ Right to “Family Life”: How Relevant Are the Principles Applied by the European Court of Human Rights?’ (2018) 25(3) *International Journal on Minority and Group Rights* 401.

⁹² See, eg, Peter J Spiro, ‘A New International Law of Citizenship’ (2011) 105(4) *American Journal of International Law* 694.

⁹³ *Mabo (No 2)* (n 56) 51, 58.

⁹⁴ *Chetcuti* (n 12) 710 [12] (Kiefel CJ, Gageler, Keane and Gleeson JJ).

⁹⁵ *Ibid* 710 [13] (Kiefel CJ, Gageler, Keane and Gleeson JJ).

the apparent rejection of key elements of the *Love* majority's reasoning may leave that exception without a principled foundation.

B *What Does Constitutional Membership Entail?*

The clearest consequence of community membership (or non-alienage) arising from *Love* is freedom from liability to removal or exclusion. It may not be much of a stretch to suggest that persons falling outside the scope of the aliens power for the reasons discussed in *Love* would likely also fall outside the scope of the immigration power, on the basis that Australia is their home,⁹⁶ although community membership in the context of the immigration power is not identical to that under discussion here. Beyond this, however, Gageler J has described the consequences of the majority's reasoning as relegating persons like the plaintiffs to a 'constitutional netherworld', where they are members of the community but lack the status of citizens.⁹⁷

Potentially the most pressing issue concerns the distribution of political rights, currently attributed on the basis of statutory citizenship. Gordon and Edelman JJ's reasoning, concerning popular sovereignty, most clearly suggests that political rights would follow constitutional membership. References to 'the people of Australia' invite consideration of constitutional principles of popular representation, preventing exclusion from the franchise otherwise than on the basis of 'substantial reasons' bearing a 'rational connection with the identification of community membership or with the capacity to exercise free choice'.⁹⁸ Bell and Nettle JJ's approaches do not so obviously link constitutional community membership to questions of popular representation, but nor do they rule it out. As for the minority in *Love*, it remains to be seen how the commitment to a citizen/alien binary will be applied in light of the majority's conclusion that there exist constitutional community members.

The relationship between constitutional membership and statutory citizenship also remains unresolved: are there people who are statutory citizens, but not constitutional community members (because they lack the relevant connection to territory) or is the effect of formal naturalisation to bring such persons within the constitutional category? If some citizens are not constitutional community members, are they liable to redefinition as aliens by Commonwealth legislation? The result in *Ex parte Ame* seems to suggest this possibility, but in that case Kirby J at least went to significant lengths to explain how the legislative transformation of citizens to aliens, and the consequent loss of constitutional protections, were only possible in the context of external territories.⁹⁹ The result in that case may be more satisfactorily confined to its own particular facts following *Love*, on the basis that the territory to which affected persons had the relevant connection had ceased to be part of the Commonwealth of Australia.

⁹⁶ See *Potter v Minahan* (n 5).

⁹⁷ *Love* (n 1) 210 [131] (Gageler J).

⁹⁸ *Roach v Electoral Commissioner* (2007) 233 CLR 162, 174–5 [7]–[8] (Gleeson CJ). See also *Rowe v Electoral Commissioner* (2010) 243 CLR 1.

⁹⁹ *Ex parte Ame* (n 22) 477–8 [99]–[101] (Kirby J). The rest of the court took a somewhat more ambiguous position: at 458–9 [33]–[35] (Gleeson CJ, McHugh, Gummow, Hayne, Callinan and Heydon JJ).

Recent comments by the Full Court of the Federal Court of Australia have highlighted the significance, albeit in a different context, of the acquisition of citizenship ‘in terms of unrestricted membership of the Australian community as a whole’,¹⁰⁰ where the community in question ‘may be understood to be the people referred to in the *Constitution*’.¹⁰¹ Most recently, Edelman J in *Chetcuti* has suggested that while a constitutional non-alien may subsequently become an alien by ‘the application of the *Constitution* to new political and social facts and circumstances’,¹⁰² and while the grant of statutory citizenship is not a ‘constitutional ratchet’ capable of preventing such changes,¹⁰³ norms of citizenship have nonetheless taken on a significant influence on the scope of the aliens power.¹⁰⁴ This leaves open the possibility that, applying a contemporary concept of community membership, all citizens may nonetheless be found to fall beyond the scope of the aliens power.

V Conclusion

The High Court’s decision in *Love* gives substance to previously elusive limits on the aliens power under s 51(xix) of the *Constitution*, while leaving significant questions to be answered by future litigation. The minority’s valid-criterion-of-alienage approach, which would have given the Commonwealth Parliament an extremely broad power to exclude persons from the Australian community, was rejected in favour of a concept of constitutional community membership. How broad this concept will prove to be, and what consequences and rights will flow from it, remain to be seen. Indeed, there are already suggestions from the joint judgment in *Chetcuti* that key elements of the majority’s reasoning in *Love* may be ignored following recent changes to the composition of the High Court. Nonetheless, the various approaches taken by the majority in *Love* provide some reason to believe that any future developments in this space will likely involve a territorial principle of membership, rather than the more problematic concept of allegiance.

¹⁰⁰ *Minister for Home Affairs v Lee* [2021] FCAFC 89, [81] (Logan, Kerr and Banks-Smith JJ).

¹⁰¹ *Ibid* [83], citing *Love* (n 1) [10] (Kiefel CJ).

¹⁰² *Chetcuti* (n 12) 719 [56].

¹⁰³ *Ibid* 722 [69]. See also at 727 [90]: ‘There is no person whose constitutional status with respect to alienage is immune from any change in facts and circumstances and is therefore indelible.’

¹⁰⁴ *Ibid* 721–3 [65], [70]–[72] (Edelman J).