

TABLE OF CONTENTS

THE 2012 JOHN BRAY ORATION

John Doyle Choosing Our Judges 223

ARTICLES

M Stuart Madden Efficiency Themes in Tort Law from Antiquity 231

Chris Finn Extrajudicial Speech and the Prejudgment Rule:
A Reply to Bartie and Gava 267

Jenny Buchan Franchising: A Honey Pot in a Bear Trap 283

Mirko Bagaric
and Theo Alexander A Rational Approach to Sentencing White-Collar
Offenders in Australia 317

Tareq Al-Tawil Damages for Breach of Contract: Compensation,
Cost of Cure and Vindication 351

Theodore Bennett Transsexualism and the Consideration of Social
Factors within Sex Identification Law 379

Michael Legg and
Dean Jordan The Australian Business Judgment Rule After
ASIC v Rich: Balancing Director Authority and
Accountability 403

Mitchell Landrigan Can the Implied Freedom of Political Discourse
Apply to Speech by or about Religious Leaders? 427

CASE NOTE

Emily Bell *New South Wales v Kable* (2013) 298 ALR 144 459

BOOK REVIEWS

Andrew Lynch *Judicialization of Politics: The Interplay of
Institutional Structure, Legal Doctrine and
Politics on the High Court of Australia* 465

Sheelagh McCracken *Australian Personal Property Securities Law* 473

SUBMISSION OF MANUSCRIPTS 476

Adelaide Law Review Association
Adelaide Law School
The University of Adelaide
Adelaide SA 5005
AUSTRALIA

Adelaide Law Review 2013

Vol 34 No 2



THE UNIVERSITY OF ADELAIDE
ADELAIDE LAW REVIEW
ASSOCIATION

ADELAIDE LAW REVIEW

VOLUME 34 NUMBER 2



THE UNIVERSITY OF ADELAIDE
ADELAIDE LAW REVIEW ASSOCIATION

ADVISORY BOARD

Emeritus Professor W R Cornish
Emeritus Herchel Smith Professor of Intellectual Property Law
University of Cambridge
United Kingdom

Professor J R Crawford
Whewell Professor of International Law
University of Cambridge
United Kingdom

Emeritus Professor M J Detmold
Adelaide Law School
The University of Adelaide
South Australia

The Hon Professor J J Doyle
Former Chief Justice
Supreme Court of South Australia

Emeritus Professor R Graycar
Sydney Law School
The University of Sydney
New South Wales

Professor J V Orth
William Rand Kenan Jr Professor of Law
The University of North Carolina at Chapel Hill
United States of America

Emeritus Professor I Shearer
Sydney Law School
The University of Sydney
New South Wales

Professor J M Williams
Adelaide Law School
The University of Adelaide
South Australia

ADELAIDE LAW REVIEW

Editors

Paul Babie John Gava

Book Review Editor

Gabrielle Appleby

Senior Student Editors

Zoe Irwin Richard Sletvold

Student Editors

Michael Armstrong

Emily Bell

Claire Benn

Lauren Clark

Tania Drewer

Rebecca McEwen

Michael Norris

Seb Tonkin

Katherine Varsos

Alice Wharldall

Publications Officer

Panita Hirunboot

Volume 34 Issue 2 2013

The *Adelaide Law Review* is a refereed journal that is published twice a year by the Adelaide Law Review Association of the Adelaide Law School, The University of Adelaide. A guide for the submission of manuscripts is set out at the back of this issue. Articles and other contributions for possible publication are welcomed.

Copies of the journal may be purchased, or a subscription obtained, from:

Administrative Officer
Adelaide Law Review Association
Adelaide Law School
The University of Adelaide
South Australia 5005
AUSTRALIA

e-mail: <panita.hirunboot@adelaide.edu.au>

For North America:
William S Hein & Co
1285 Main Street
Buffalo NY 14209
USA

<<http://www.law.adelaide.edu.au/review>>

This volume may be cited as:
(2013) 34 *Adelaide Law Review*
The articles in this volume are published in 2014.

ISSN 0065-1915

© Copyright is vested in the Association and, in relation to each article, in its author, 2014.

TABLE OF CONTENTS

THE 2012 JOHN BRAY ORATION

John Doyle	Choosing Our Judges	223
------------	---------------------	-----

ARTICLES

M Stuart Madden	Efficiency Themes in Tort Law from Antiquity	231
Chris Finn	Extrajudicial Speech and the Prejudgment Rule: A Reply to Bartie and Gava	267
Jenny Buchan	Franchising: A Honey Pot in a Bear Trap	283
Mirko Bagaric and Theo Alexander	A Rational Approach to Sentencing White-Collar Offenders in Australia	317
Tareq Al-Tawil	Damages for Breach of Contract: Compensation, Cost of Cure and Vindication	351
Theodore Bennett	Transsexualism and the Consideration of Social Factors within Sex Identification Law	379
Michael Legg and Dean Jordan	The Australian Business Judgment Rule After <i>ASIC v Rich</i> : Balancing Director Authority and Accountability	403
Mitchell Landrigan	Can the Implied Freedom of Political Discourse Apply to Speech by or about Religious Leaders?	427

CASE NOTE

Emily Bell	<i>New South Wales v Kable</i> (2013) 298 ALR 144	459
------------	---	-----

BOOK REVIEWS

Andrew Lynch	<i>Judicialization of Politics: The Interplay of Institutional Structure, Legal Doctrine and Politics on the High Court of Australia</i>	465
Sheelagh McCracken	<i>Australian Personal Property Securities Law</i>	473

SUBMISSION OF MANUSCRIPTS		476
----------------------------------	--	-----

THE 2012 JOHN BRAY ORATION

*The Hon John Doyle AC QC**

CHOOSING OUR JUDGES

Sometimes criticism of judicial decisions leads to proposals to change the method of selecting and appointing judges and magistrates. Sometimes proposals for change are prompted by other factors.

The arguments for change usually reflect familiar themes. Some are concerned mainly with process. The appointment process should be written down, made public and should be open to public scrutiny. A variant on this process argument is an argument that the role of the elected government should be reduced or eliminated, and that the power to appoint should be given to a commission or committee, independent of the government. Other arguments reflect a wish to change the composition of courts. When you think about it, this must be with a view to changing the decisions reached. This, however, is rarely spelt out.

As we all know, the judiciary is largely male. This is changing rapidly. The number of women appointed to the bench is increasing steadily. Some people want the composition of the judiciary to reflect the composition of South Australia's population. They are rarely specific about how far they would go in that direction. Do they mean a bench that reflects the gender balance in the community? Do they want different ethnic groupings reflected in the bench? Do they want judges from rural areas as well as from urban areas? One could go on and on. The short answer to this is that it is impossible to insist on reflecting the composition of our population, if suitability for judicial office is to remain the primary criterion for appointment.

Another argument used is that judges should be more in touch with the community. What this often means is that the judge should agree with the speaker or writer. Some people want to influence how judges decide cases. They think that judicial attitudes are too much influenced by the professional background of members of the judiciary. So they want to change the composition of the bench with a view to changing the attitudes of the judiciary, and the decisions that the judiciary make.

The effect on our judiciary and its decisions of changing the method of selection is one of the great unknowns. Advocates of changing the selection methods sometimes assume that this will change the decisions that the court makes. No-one really knows. However, putting this to one side, it is worthwhile considering whether our method of selecting judicial officers could be improved.

* Adjunct Professor, Adelaide Law School, The University of Adelaide; Chief Justice of South Australia, 1995–2012. This is an edited version of John Doyle, 'Choosing Our Judges' (Speech delivered as The 2012 John Bray Oration, The University of Adelaide, 1 September 2012).

Sometimes the debate about choosing judges results in suggestions that judges and magistrates should be elected. The argument runs along these lines: We are a democracy, judges are public officials, and therefore they should be elected. Judges should be accountable, and so the people should be able to remove them. Elections would make the judiciary more representative. The need to get elected, and the risk of not being re-elected, would make the judiciary more responsive to public opinion. Judges are elected in America, so why not here?

Elections are said to achieve democratic legitimacy, accountability, a representative bench and a responsive bench. People who promote electing judges and magistrates think that it will achieve in one hit all of the changes that they want. They draw on the American experience because America is the only democracy, so far as I am aware, that elects its judges. Before dealing with these arguments, it might be helpful if I outline how we appoint judges and magistrates in South Australia.

Throughout Australia judges and magistrates are usually appointed by the Governor, acting on the recommendation of Cabinet or of the Attorney-General. So, the real decision is made by Cabinet or by the Attorney-General. Although the formal decision is made by the Governor, the Governor has no choice but to accept the advice of the Minister or of Cabinet.

How does Cabinet make its decision? The process is usually for the Attorney-General to make a recommendation to Cabinet. The extent to which the Cabinet involves itself in the decision will vary from place to place.

How does the Attorney-General arrive at a recommendation? There is some flexibility in this. I believe that throughout Australia the practice is fairly uniform. In South Australia the Attorney-General usually consults with the Chief Justice and, if the appointment is to a court other than the Supreme Court, with the judicial head of that Court. The Attorney consults with the President of the Law Society and the President of the Bar Association. I understand that the Attorney may consult with the Shadow Attorney-General, with the Solicitor-General, the Crown Solicitor and the Director of Public Prosecutions. Then the Attorney-General is at liberty to consult with other people. I assume that these would include fellow ministers and selected people within the community.

The process is not a public one. However, a wide range of views will be obtained.

The privacy of the process is said to be important. Some people would not comment frankly on particular people under consideration if they knew that their opinions might become public. It could be unfair to a person being considered for a judicial appointment if adverse comments about that person were to become public. I recognise the concerns of those who object to the privacy of the process but there is something to be said for it.

I emphasise that the judicial head of the relevant court does not have the final say. By convention the Attorney-General consults with them, but need do no more than consult. The expectation is that competence and personal integrity are the most important

criteria. If two candidates were in these respects equal, it is legitimate for the Attorney to make a choice on the basis of, for example, gender or background. The process has worked well, even though it is informal and not controlled by law, and not conducted in the public gaze. At no stage is a list of possible appointees published.

There have been suggestions for change.

One is that a statutory commission should be established with the power to make appointments. It might include senior judicial representatives, the Attorney-General, and representatives of the public. Another occasional suggestion is that such a commission be established to draw up a list of persons, with the government still making the choice, but being obliged to choose a person from the list, or, as an alternative, to give reasons if a person not on the list is chosen.

My opinion is that each of these proposals is a reasonable one, but I am not sure that either of them would amount to a significant improvement over the existing system. For what it is worth, I favour the second one of these. I believe that the government has a legitimate role in the appointment of the judiciary, and giving the government a role adds a democratic element to the process.

I return to the question of elections. What is the position in America? In America Federal judges — judges appointed by the United States government — are selected by the executive as occurs here. My belief is that the process is more openly political and more public than it is here. Political factors, and the attitudes of candidates, are canvassed much more openly than is the case in Australia. This probably reflects, in part, the fact that the role of judges in interpreting constitutional rights gives them a greater influence on political decision making.

A distinctive feature of the United States process is the process of confirmation hearings before a Senate committee. At these hearings candidates are questioned about their approach to issues, and their past can be raked over, as has happened on more than one occasion. Not all candidates survive. It seems clear that the process is not a pleasant one for those who have to go through it.

It is only at the state level that judges are elected in America. Articles that I have read state that judicial elections have been used in America since the middle of the nineteenth century. In 39 states, some or all judges are elected.¹ About 80 per cent of all state judges face election at some point in their career.² I put it this way because in some states you have to go through an election to be appointed at all, in other states you can be appointed to the bench but have to go through an election to secure a further term.³ The importance of elected judges emerges from the fact that apparently over 90 per cent of all court business in America occurs in state trial courts.⁴

¹ Justice Margaret H Marshall, 'Dangerous Talk, Dangerous Silence: Free Speech, Judicial Independence and the Rule of Law' (2002) 24 *Sydney Law Review* 455, 464.

² *Ibid.*

³ *Ibid.*

⁴ *Ibid.*

The Chief Justice of Wisconsin, Shirley Abrahamson, is an elected judge. She supports judicial elections. This is what she said, having stood successfully at three elections, each of which she said was hotly contested:

The last campaign involved such lofty issues as the appropriateness of my sponsoring a staff aerobic class in the courtroom after hours, my decision to hang a portrait of the first woman to be admitted to the Wisconsin Supreme Court bar, and the removal of computer games from justices' computers. I was also the target of television ads challenging my votes in specific cases involving a school locker search, a *Terry* stop, and the sexual predator law. More than \$1 million was raised by both candidates, and the election set a record for Wisconsin judicial campaign spending. The most fun thing about the race was winning it.⁵

You may be interested to know that in America the election of judges is opposed by many, including organisations representing the legal profession, and by many Federal judges and organisations associated with the Federal judiciary.

Chief Justice Abrahamson put several arguments in support of electing judges, which I will mention briefly. First, she makes the point that there is no perfect system for selecting judges. Every system has its strengths and weaknesses.⁶ I agree. She says that the elective system can be an educational experience for both the judges and the electorate.⁷ I must say I doubt that, but I am not really in a position to differ. She makes the point that in a democracy, an unelected judge cannot justify overturning legislation adopted by a democratically elected legislature.⁸ I firmly disagree with this proposition. She says that no study has proved that elected judges perform poorly compared with appointed judges.⁹ She says that elections are not a threat to judicial independence, as long as judges are true to their judicial oath.¹⁰

Chief Justice Abrahamson, Roy Schotland¹¹ and James Sample¹² make some valid points. The first point is that in an elective system much more information is out in the open than is the case in an appointed system.¹³ Both systems (that is, our system and the American system) are said to be partly politicised, but one of them, the

⁵ Shirley S Abrahamson (Chief Justice), 'The Ballot and the Bench' (2001) 76 *New York University Law Review* 973, 975.

⁶ *Ibid* 976.

⁷ *Ibid* 977.

⁸ *Ibid* 979.

⁹ *Ibid*.

¹⁰ *Ibid* 984.

¹¹ Roy A Schotland, 'New Challenges to States Judicial Selection' (2007) 95 *Georgetown Law Journal* 1077.

¹² James J Sample, 'Court Reform Enters the Post Caperton Era' (2010) 58 *Drake Law Review* 782.

¹³ Abrahamson, above n 5, 993–94.

elective system is open about it.¹⁴ Another point made is that in an elected system greater diversity can be attained because people who would otherwise be on the outer and unlikely to be considered, can run for election and obtain the position that they would otherwise never obtain.¹⁵

I am opposed to the election of judges. It may work satisfactorily in America, although as I will tell you shortly, things are happening in America that cause supporters of elections to have some doubts. But even if it can work satisfactorily, I do not believe we could graft the election of judges onto our system. We need to remember that in America the election of judges occurs in a society in which many officials, whom we would appoint, are elected to office, and there is now a long tradition of judicial elections.

This is why I think elections are inappropriate.

A judge is appointed on the basis of the judge's professional skills, personal integrity and independence of mind. I am not confident that a process of election would result in appointments being made on the basis that these are the decisive criteria. In an election other factors are likely to play a greater part. Factors that might become relevant are the ability to raise campaign funds, being more photogenic and handling the media better. No doubt you can think of other things that would be relevant to an election. Also, imagine the outcry if it was said that we should elect people to fill senior surgical positions in the public hospitals. Our tradition is that professional positions are filled by appointment.

Next, there is the question of whether the best people would be willing to stand for election. I suggest that a number would not.

A particular point I make is the importance of the independence of the judiciary. Experience suggests that if there are elections, a candidate for judicial officer will have made promises or indicated attitudes to campaign supporters or to voters, and will have been helped financially and in other ways. Is there not a danger that successful candidates will come to the bench owing obligations to particular interest groups? This would be unsatisfactory in my opinion. Also, would not candidates be seen as representing particular interest groups, at least on particular issues raised in an election campaign? That would be unsatisfactory. Judicial independence would be compromised.

Imagine that you had the misfortune to appear before a judge who was elected on a platform of being tougher on crime, and who is shortly to face re-election. The circumstances look rather suspicious, and you are pleading not guilty. Would you be confident that such a judge would try your case and sentence you, uninfluenced by the fact that, with an election looming, an acquittal might be seized on by the judge's opponents as indicating that the judge is 'soft on crime'?

¹⁴ Schotland, above n 11, 1086.

¹⁵ Ibid 1090.

In the paper to which I referred, Chief Justice Marshall of Massachusetts acknowledges that there are some recent developments that cause concern. She says that since the late 1980s judicial elections, which had previously been low key, have become ‘dog fights of a high order’.¹⁶ This is because various interest groups have realised the significance of the judiciary in relation to the interests that they promote. They have begun to support and oppose particular judges standing for election by reference to their attitude on the issue that concerns them. Groups she mentions are trial lawyers, pro-business groups, pro-abortion and pro-life organisations. She says that in 2000, in the five States with the most hotly contested judicial elections, almost \$35m was raised by the candidates. To me, as to her, that is a troubling fact.¹⁷

The most recent figures I have are in Schotland’s article. He says that in 2004 there were 49 judicial seats up for election, not counting retention elections. Candidates raised \$46.8 million, and there was a further \$12 million spent on television advertisements funded directly, and not by candidates. In nine of the 16 States where there were elections, new records were set for expenditure.¹⁸ Television advertisements were used in all of the States. Chief Justice Marshall also says that ‘pugnacious advertisements’ by candidates are cheapening judicial office.¹⁹ In his article Schotland gives some interesting examples. Each of them is a little dated, so perhaps things have improved.

First of all, in California in 1984 an advertisement boasted ‘[s]ent more criminals — rapists, murderers, felons — to prison than any other Judge in Contra Costa County history’.²⁰ Another from 1994 announced ‘[o]ver 90 per cent convicted criminals sentenced ... prison commitment rate is more than twice the State average’.²¹ Finally, Chief Justice Marshall mentions that in a recent case the United States Supreme Court held unconstitutional a common state statute, which prohibited candidates for judicial office from ‘announcing his or her views on disputed legal or political issues’.²² That means that candidates can now be driven to indicate their views on issues likely to come before them. They can no longer say that the law forbids them to do so. To me it is most undesirable for that to happen. First of all, it can give rise to a serious concern about the issue of independence. If I am elected with the backing of a pro-life lobby, having stated that I am strongly opposed to abortion, what happens if a person comes before me charged with an offence of that kind?

All judges have opinions, and often they are opinions about issues that come before the court. By and large people accept that judges can put their personal opinions and attitudes to one side, and judge a case fairly. But if a judge has made a public

¹⁶ Marshall, above n 1, 465.

¹⁷ Ibid 465.

¹⁸ Schotland, above n 12, 1080.

¹⁹ Marshall, above n 1, 465.

²⁰ Schotland, above n 11, 1099.

²¹ Ibid.

²² Marshall, above n 1, 466.

commitment in the process of getting elected, it is too much to think that people will accept that the judge will not be influenced by that commitment. Of its very nature, it is a commitment relating to how the judge will perform the judicial office.

Chief Justice Abrahamson also acknowledges some causes for concern. The first of these is a public perception that judges are influenced by campaign contributions. As she rightly says, on a matter like this perception is as important as reality.²³ If voters think that donors are calling the judicial tune, confidence in the judiciary must be undermined.

Second, she remarks that judicial elections have become very expensive. Candidates are putting their own money into their elections.²⁴ She makes the telling point that a position on the bench might become limited to the wealthy, if the cost of elections reaches the point at which only the wealthy can afford to compete.²⁵ That would be a disastrous outcome from a method of appointment based on the theory of democracy. Costs might close off a judicial career except to those who are wealthy or who can attract significant financial support. How can this be said to promote diversity? If anything, it narrows the pool of candidates.

So you can see that electing judges, while at first blush it might seem attractive, raises a number of problems.

Where do we go from here?

My opinion is that in Australia the only realistic possibility for change is the establishment of a committee or a commission that has the power to select a list of candidates from which the government must make its choice. I doubt whether any Australian government would go further than that. The composition of a selection committee or commission, and the vesting of the power of appointments would be contentious. The debate about who chooses judges would simply shift from debate about the government making the choice, to debate about a committee making the choice. There would not be any net gain.

I am confident that, over time, we will see an increasing number of women on the bench. The gender balance is changing, and will do so whatever method of appointment we choose. I am also confident that the background of the judiciary is changing, reflecting the change in the intake in our universities and law schools.

I can assure you that the judiciary is a diverse group of people. Our backgrounds vary greatly. We are not mainly 'silver spooners'. The attitudes of the judiciary on social issues are very mixed. However, it is inevitable that we reflect our professional background. That should not be seen as a bad thing. I want my doctor's approach to treating me to reflect his professional background. Unfortunately, in some quarters

²³ Abrahamson, above n 5, 995.

²⁴ *Ibid.*

²⁵ *Ibid.*

there is an impression that judges are all elderly men who are deeply conservative in their approach, except when they make an adventurous decision of the kind the High Court sometimes makes, and then the problem is that they are not conservative enough. We cannot win.

Time will tell, but for these reasons I suggest that if you want to elect a judge, go to America to live.

EFFICIENCY THEMES IN TORT LAW FROM ANTIQUITY

ABSTRACT

As human societies developed, a bedrock necessity was the identification of expectations and norms that protected individuals and families from wrongful injury, property damage, and takings. Written law, dating to the Babylonian codes and early Hebrew law, emphasized congruent themes. Such law protected groups and individuals from physical or financial insult, depredation of the just deserts of labor, interference with the means of individual livelihood, and distortion of the fair distribution of wealth.

Hellenic philosophers assessed the goals of society as being the protection of persons and property from wrongful harm, protection of the individual's means of survival, discouragement of self-aggrandizement, and the elevation of individual knowledge that would carry forward and perfect such principles. Roman law was replete with proscriptions of forced takings and unjust enrichment, and went so far as to include rules for ex ante contract-based resolution of potential disagreement. Unwritten customary law within the Western world and beyond perpetuated these tenets, based at once in morality and aversion to wasteful behavior.

In addition to the corrective justice-morality underpinnings of the law governing civil wrongs, or torts, the common law has nurtured rules implicating economic and efficiency themes. Efficiency themes enjoy a conspicuous place in modern tort analysis: from the risk-utility analysis and implicit social cost evaluations of numerous common law courts in accident cases, to the translation of the negligence formula of Judge Learned Hand into a basic efficiency model, to the increasing number of judicial opinions that rely explicitly upon economic analysis.

* Former Distinguished Professor of Law, Pace University School of Law. Email address: mstuartmadden@yahoo.com.

I INTRODUCTION

Tort law represents a society's revealed truth as to the behaviors it wishes to encourage and the behaviors it wishes to discourage.¹ From causes of action for the simple tort of battery to the more elegant tortious interference with prospective advantage, the manner in which individuals or groups can injure another seems limitless. Despite the amplitude of interests protected by tort law, from its earliest exercise in prehistoric groups up to its modern implementation, there have existed a finite number of goals of tort law, whether the 'law' referred to be an unwritten norm, a judicial decision, or a modern statute.

There is general agreement that these objectives, however imperfectly accomplished, include: (1) returning the party who has suffered a loss to the position he enjoyed before the wrongful activity; (2) requiring the wrongdoer to disgorge the monetary or imputed benefit derived from his actions; and (3) by the remedy meted out, and by its example, deterring the wrongdoer and others in a similar situation from engaging in the same wrongful and injurious pursuit. Another manner of describing tort goals has been to order them as serving either goals of (4) 'corrective justice' and 'morality'; or (5) 'economic efficiency and deterrence'.

Nonetheless, the philosophical underpinnings of tort rules is not the syncretic hotchpotch it may initially appear, even though aligning the rules exclusively with any one of the five goals requires some ungainly packaging. Each of the five themes described actually furthers the other four. This is to say, a remedy that focuses on corrective justice will serve simultaneously the goals of disgorging the wrongdoer of his unjust enrichment, morality, efficiency, deterrence, and so on. More specifically, the goal of returning the injured party ('plaintiff') to the status quo ante, the objective most closely associated with corrective justice, is ordinarily reached by a decree ordering the wrongdoer to pay to the plaintiff the money equivalent of what the plaintiff lost, with damages calculated in this way operating as an inexact surrogate for what the wrongdoer gained, actually or by imputation, by perpetrating the wrong. Further, while the wrongdoer's forfeiture of his gain not only provides corrective justice for the plaintiff, it also punishes the wrongdoer for failing to achieve the plaintiff's ex ante approval of the transaction — an omission deemed by economic theorists to be inefficient.² Although it is not surprising that many suggest that tort rules and remedies aligned with economic and efficiency models provide optimal

¹ There will be some rarified instances of behavior for which tort law would not encourage elimination, such as abnormally dangerous activities ranging from construction blasting to aerial application of pesticides, but instead may wish to modify or limit, and in any event, to assign strict liability. See *Indiana Harbor Belt Railroad Co v American Cyanamid Co*, 916 F 2d 1174, 1177 (7th Cir, 1990); see generally Steven Shavell, 'Strict Liability versus Negligence' (1980) 9 *Journal of Legal Studies* 1.

² When the loss is personal injury or property damage, a rough estimation of this inefficiency (or waste) may often be the combined amount of the claimant's economic and non-economic damages. Of course the themes of punishment and deterrence are but the flip side of the goal of creating an incentive for efficient behavior. As suggested by Professors David W Barnes and Lynn A Stout, '[t]ort law may be viewed as a

deterrence for civil, tort-type wrongs, the following discussion confirms that the tort rules recognized by the corrective justice-morality school also deter in measurable ways. Indeed, in the inexact taxonomy employed by tort scholars, there are so many instances of overlap between tort goals that are claimed to serve corrective justice-morality, but simultaneously serve goals of efficiency and deterrence, that the legal pragmatist would be tempted to characterize them as functionally equivalent.³ Even conceding the absence of neatness in any attempt at categorization, the division of tort goals along these or similar lines is nevertheless illuminating and predictive.

Tort law is a model of social expectations, and these social expectations are at once moral and economically efficient. The goal of this article is not to elevate the consideration of economic objectives over those of corrective justice and morality, or the reverse. Rather, the author seeks to survey the ample social and legal record, revealing that, over history, written or unwritten rules pertaining to civil wrongs have cleaved as readily to an ethos of efficiency as such rules have promoted goals of corrective justice and morality.

This efficiency norm has organizing principles of waste avoidance, the protection of persons and their property from injury and wrongful appropriation, the preservation of the integrity of individual or collective possessions or prerogatives from wrongful interference, and the prudent marshaling of limited resources. Even when the corrective justice-morality objective of any tort rule may appear on its face to eclipse any efficiency underpinnings, subtle economic themes of efficiency and deterrence can be recognized in almost all tort-type customs, expectations, and rules. This article will demonstrate that the parallel and harmonious impetus for almost all of what we call tort law today can be found in principles of corrective justice and economic efficiency, and that individually, corrective justice and efficiency are each necessary, but neither sufficient, premises for explaining past and current tort rules.

In any evaluation, social or scientific, as it has evolved during the period of written history, there are inevitable gaps in the record. As regards this socio-legal history, the potential for analytical error is compounded by the difficulties legal scholars and historians confront in reading the historical record within the only context that may reveal it reliably, which is to say, the cultural and political circumstances of its origins. Moreover, as to pre-history, no more than a small part of the record of the earliest human societies may ever be scientifically reconstructed, because of natural loss or later human meddling. Forever inaccessible are countless ancient remnants that might suggest the social norms employed to make group decisions based on what behaviors would bring collective benefit and what would not.

system of rules designed to maximize wealth by allocating risks so as to minimize the costs associated with engaging in daily activities': David W Barnes and Lynn A Stout, *Cases and Materials on Law and Economics* (West Publishing, 1992) 82.

³ Put another way, both corrective justice and efficiency principles must be regarded as 'true' in that they hold significant, albeit non-exclusive, predictive value in anticipating the development of tort law. See M Stuart Madden, 'Selected Federal Tort Reform and Restatement Proposals Through the Lenses of Corrective Justice and Efficiency' (1998) 32 *Georgia Law Review* 1017, nn 297–8 and accompanying text.

The adoption of durable writing or imagery accelerated our modern understanding of ancient legal rules. The discovery and translation of the first integrated legal codes from the sites that were within ancient Babylonia, codifications of what was surely the customary law that preceded it, provided the first written evidence of regularized standards for individual behavior, identification of civil wrongs, and the remedies for such wrongs.

Hobbes described survival as man's strongest moral imperative, and that all pursuit of justice and right is founded in man's rational pursuit of self-preservation.⁴ Experts are of one view that the success and survival of early social groupings bore a more or less exact correlation to their adoption of rules that furthered advancement of knowledge, material comfort, and economic stability. Achievement of these attributes would, from pre-history onward, be characterized as 'good'.⁵ It follows that early family clans, and the tribes and ever-larger social aggregations that would follow, shared one sentiment: to pursue such 'good' for their members.⁶ Philosophers have disagreed as to whether man in his natural state was innately 'good', but any original impulse for good stood no meaningful chance for survival as human concentrations grew and evolved. Group order and expectations in the form of norms, and the subscription to such norms by individuals and families, became necessary for communal survival. It will be seen that at its core, tort law, together with its unwritten normative antecedents, bears witness to the fundamental social need for self-limitation. To the sociologist Emile Durkheim, the peaceful process of society has always depended on the individual's submission to inhibitions of or restrictions on personal 'inclinations and instincts'. To Durkheim, 'social life would be impossible'⁷ without general subscription to such limitations, and this would hold true whether the 'venerable respect' tendered to a collective 'moral authority' is faith-dependent or not.⁸ And so by necessity, social groups developed expectations, norms, customs, and, eventually, laws that encouraged behaviors that contributed to the common good and economic success of the community; and

⁴ Thomas Hobbes, *Leviathan* (Bobbs-Merrill, first published 1651, 1958 ed), discussed in Francis Fukuyama, *The End of History and the Last Man* (Free Press, 1992) 156; Leo Strauss, *Natural Rights and History* (University of Chicago Press, 1953) 187–8.

⁵ See generally Robert Redfield, 'Maine's Ancient Law in the Light of Primitive Societies' (1950) 3 *Western Political Quarterly* 574, 586–9, in which Redfield writes of primitive societies: 'economic systems are imbedded in social relations. Men work and manufacture not for motives of gain. They tend to work because working is part of the good life'.

⁶ Erich Fromm confirmed 'the strength of the strivings for happiness and health that are part of the natural equipment of man'. Erich Fromm, *Man for Himself: An Inquiry Into the Psychology of Ethics* (Fawcett Publications, 1947) viii. By 'members' is meant the collective. As Maine observed: 'Ancient Law ... knows next to nothing of Individuals. It is concerned not with Individuals, but with Families, not with single human beings but groups': Henry Sumner Maine, *Ancient Law: Its Connection With the Early History of Society and Its Relation to Modern Ideas* (John Murray, 1861) 229.

⁷ Emile Durkheim, *The Elementary Forms of Religious Life* (Free Press, first published 1912, 1965 ed) 237.

⁸ Subscription to this logic will not be found among those who believe that faith is merely an attempt to put some structure on the chaos that surrounds the human experience.

discouraged individualistic pursuit of personal aggrandizement to the extent that it involved disavowal of community responsibility.

Accordingly, human experience of the ages has demonstrated that man as a social animal has turned almost invariably to structures and norms consistent with defined and enforced standards of 'good' as would further the innate and overarching instinct for individual and group survival. By virtue of this ascendant sentiment of most societies of all historical epochs to attain both group and individual 'good', the collective conclusions as to what constitutes 'good' evolved gradually to this: what is 'good' has always been, as it is today, what is just, moral and equitable.⁹ Encouragement of 'good' conduct has been logically accompanied by discouragement of 'bad' conduct, which is to say, behavior considered to be unjust, immoral, or inequitable. And all such systems, save the brashest of totalitarian societies, have included standards by which a person might seek the correction of or compensation for harm caused by the wrongful acts of another. Initially established as practices, then as norms and customs, and eventually as law, evolving social strictures would operate to either cabin or punish the behaviors of those succumbing to the seemingly irresistible human appetite for bad, wrongful, and harmful behavior.

In this sense, tort law past and present, has operated as the societal super-ego, a generally subscribed-to social compact in which most persons rein in such impulses as might lead them to trammel the protected rights of others, inasmuch as the norms of tort law require rectification operating post hoc to restore the wronged person to the position previously enjoyed. This restoration may be perfect, such as when it is in the form of returning goods where there has been a trespass to chattels and there has been no diminution in value, or when there has been a misappropriation. Or it may be imperfect, such as in settings involving a wrongful physical injury, as to which rectification in the form of money can never truly restore the injured party to the status quo ante. As suggested initially, whatever the corrective justice limitations of money damages, they do serve other objectives identified with tort law, which include deterrence of the same or similar conduct by the actor or others similarly situated.¹⁰ Money damages also, in an economic sense, command a transfer of wealth that achieves a figurative rectification of the wrongdoer's 'forced taking' of the injured party's bodily integrity. The money damages also, at least conceptually, deprive the wrongdoer of the 'unjust enrichment' achieved by creating a tear in the fabric of consensual or contract-based social interaction.

⁹ Conceding that Socrates wrote from beyond the spheres of governing power, it is telling that Socrates' ethics are suffused with the goal of avoiding doing harm, and with the argument that a principal marker of 'justice' is the simple 'returning what was owed': Anthony Gottlieb, *The Dream of Reason: A History of Philosophy from the Greeks to the Renaissance* (Norton, 2000) 164.

¹⁰ 'The true explanation of the reference of liability to a moral standard ... is not that it is for the purpose of improving men's hearts, but that it is to give a man a fair chance to avoid doing the harm before he is held responsible for it. It is intended to reconcile the policy of letting accidents lie where they fall, and the reasonable freedom of others with the protection of the individual from injury': Oliver Wendell Holmes, *The Common Law* (Belknap Press of Harvard University Press, 1963) 115.

II ECONOMIC IMPERATIVES IN EARLY SOCIAL GROUPINGS

A Generally

The raw and primal imperative of simple human survival has required of each successful community the ordered pursuit of wellbeing for its members, necessarily including standards to discourage or interdict activity that interrupted or compromised pursuit of a social order consistent therewith. Interdependent with such overarching needs, the norms or apparatus of ‘justice’ and ‘morality’ too would sensibly harmonize with the collective pursuit of economic stability, growth, and the elevation of human knowledge. Retaining a focus on the three goals of elevation of human knowledge, material comfort, and economic stability, it follows that within the context of pre-history, of particular pertinence to the furtherance of each goal was the creation and preservation of group circumstances in which persons could expect to live peaceably without physical injury at the hands of others. It also was expected that the community would provide congruent protection against wrongful taking or damage of the property justly acquired by its members. It was collectively thought necessary that man would gradually impose on his groups, and eventually civilizations and states, norms and rules that served to protect the personal physical autonomy and security of group members, and also protect their belongings, against wrongful interference. The group visualization of these norms, and their progressive imposition, would assume the aura of inevitability, and the gravitas of a cultural imperative. For successful social groupings, principal among such norms was the expectation there would be some form of remediation for an impermissible intrusion on physical or property interests, including common property rights.¹¹ And, finally, along this line of civilizing thought, the ideation of society was that the burden of an avoidable harm ought not rest with the innocent victim but rather with the wrongdoer.

B A Pre-Symbolic Scenario

At some distant time in Africa, the birthplace of modern man, homo sapiens formed family-based social groups or clans. From the time of early family groupings to the development of ever-more complex communities, all successful human gatherings developed work specializations inter se.¹² For example, a group depending on fishing for its sustenance would need individuals to prepare nets or baskets for the catch. Others in the group would dedicate themselves to the actual fishing, and travel to the water source with, let us say, spherical fishing baskets that contained a hole on one side that lured fish to the shade. Swift retrieval of the basket would catch the fish and provide food for the community. Naturally, the entire community would not survive if the actual fishing specialists arrogated to themselves the catch, and so there

¹¹ Dennis Lloyd, *The Idea of Law* (Penguin Books, 1976) 49, referencing, inter alia, Mosaic law.

¹² As Darwin pointed out for flora and fauna, and as Durkheim noted in the case of human societies, an increase in numbers when area is held constant (ie an increase in density) tends to produce differentiation and specialization, as only in this way can the area support increased numbers.

developed norms of allocative efficiency, a so-called ‘generosity’ norm, that would ensure that all in the community, including infants and the aged, would be provided for adequately.¹³ This allocation of goods constituted a corpuscular prototype of efficiency-based exchange of goods that recognized duties owed by the community to its individuals, duties owed by community members to others, and the common interest in non-wasteful behavior that would characterize all societies to follow.

This economic cooperation characteristic of primitive communities was the antithesis of economic self-interest, and understandably, Karl Polanyi writes that in tribal society, ‘[the individual’s economic interest is rarely paramount, for the community keeps all its members from starving unless it is itself borne down by catastrophe]’.¹⁴ Thus in the circumstances of tribal society, a member’s exclusive pursuit of economic self-interest was itself contrary to the economic survival of the group. Early task assignment and economic differentiation within a clan or small social group required, by ‘code of honor’ or ‘generosity’, recognition that each member of the community served the whole. From the earliest hunting and gathering communities to the later agricultural groupings, task allocation was accompanied by mutual expectations that the bounty in food or materials gathered by one group would be shared with the others. The others would include, nonexclusively, the homemakers, children, and the elderly. For the vital hunting population to forsake its obligation to return from the hunt with food to share with the family, clan, or tribe would sabotage the very existence of the social group. Failure to share with the homemaker and the children would bring about the speedy end of the bloodline. As to elders, with some exceptions, tribal groups recognized that the aged acted not only as secondary caregivers but also as essential repositories of the group’s oral history and traditions.

In time, with the increase in population and in the course of the proved northward migration of many human groups,¹⁵ early man found that the working norms for family, clan and single community survival would be taxed by contact with other families or groups. For an untold time, the response of the principal family was simply that of preserving territorial integrity, familial safety, or both. An intruder would be frightened away, or if necessary, beaten or killed. If the intruder or his group prevailed in any contest, the principal family, with its injured or killed, would abdicate its territory.

¹³ ‘For example, the [primitive] Australian hunter who kills a wild animal is expected to give one certain part of it to his elder brother, other parts to his younger brother and still other parts of the animal to defined relatives. He does this knowing that [the other brothers] will make a corresponding distribution of meat to him’: Robert Redfield, ‘Maine’s Ancient Law in the Light of Primitive Societies’ in J C Smith and David N Weisstub, *The Western Idea of Law* (Butterworths, 1983) 81.

¹⁴ George Dalton (ed), *Primitive, Archaic and Modern Economies: Essays of Karl Polanyi* (Beacon Press, 1968) 7.

¹⁵ Such extraordinary migrations as would take man out of Africa and eventually permit his species’ dispersal throughout all but one continent was facilitated by his evolved ability to walk on two feet, to travel long distances, and to carry objects and infants: J M Roberts, *The New History of the World* (Oxford University Press, 4th ed, 2003) 5.

In a succession of discrete and unidentifiable moments, this motif would change. Increased populations, changes in climate that made one area more hospitable than another, or migratory patterns of available prey, made contact with other groups more frequent. A group had essentially two choices. They might preserve their reflexive and potentially mortal repulsion of competition. However, losses in injuries or death suffered in non-cooperative contact with other groups might have stimulated a group's conclusion that preservation of pristine territorial integrity was perhaps a pearl of too great a price. And so, alternatively, their response to other communities might begin to partake of peaceable aspects. Non-combative resolution of intra-familial allocative tensions might have served as a model for introduction of cooperative behavior in interfamilial or inter-clan matters. Cooperation would lessen or eliminate the enormous waste and cost of violent response to intrusion.

Perhaps at the instigation a group elder, families and tribes eventually developed behaviors and expectations that could coexist within the context of available resources in such ways as to achieve a tenable resource-based economic stasis.¹⁶ Should, for example, our hypothesized fishing community come into contact with a hunting community, the sharing of territory, and perhaps even barter, might well become recognized for its very significant benefit in reducing the group's loss of its ablest members to combat, and thus become a common ideal or norm.

Historians have recognized the similar options presented to later agricultural communities, with the permissible inference of the peaceable and efficient resolution of such options. In the description of J M Roberts:

As the population rose, more land was taken to grow food. Sooner or later men of different villages would have to come face to face with others intent on reclaiming marsh which had previously separated them from one another ... There was a choice: to fight or to cooperate ... Somewhere along the line it made sense for men to band together in bigger units than hitherto for self-protection and management of the environment.¹⁷

Of necessity, the norms developed within such larger social groups reflected the wisdom of not only *ex ante* resource allocation but also of strictures intended to discourage disruption of such distribution by forced takings or otherwise.

The above hypothetical, yet historically realistic, scenario offers our first chance to measure highly plausible human behavior, and attendant norms, by a yardstick of human economic efficiency. Although multiple economic models are available, one that seems well-suited is that propounded by Vilfredo Pareto in the early 1900s. The

¹⁶ Of course the genetic significance of intergroup coexistence is inestimable, but would, in any event be unknown to early man. Early intimations of the risks of sustained familial interbreeding might well be manifest in the development of the incest taboos.

¹⁷ Roberts, above n 15, 49–50. These early incentives toward political and economic cooperation weigh in against the more pessimistic vision of Garrett Hardin. See Garrett Hardin, 'The Tragedy of the Commons' (1968) 162 *Science* 1243, 1244, arguing that 'ruin is the destination toward which all men rush, each pursuing his own best interest'.

Pareto analysis imagines a setting in which all goods have been previously allocated, and permits an evaluation of different approaches to reallocation of such goods. A reallocation that left one or more individuals better off, but no one worse off, would be considered a Pareto Superior change.¹⁸ Even better, from a wealth-maximization perspective, is a result in which, with the reallocation of goods or resources, all affected parties are better off — a result described as Pareto Optimal or Pareto Efficient.¹⁹

Applying the Pareto approach to early man's described movement away from territorial combat to gradually more peaceable distribution of land and other resources presents this question: is such rational cooperation efficient? A syllogism posed in a coarse correlation between competition and efficiency may be, on these facts, misleading. That syllogism would go: competition is, generally speaking, efficient. The antithesis of competition is cooperation. Therefore, cooperation is inefficient. However, in the example given earlier, rational cooperation between early human social groups regarding the sharing of limited land resources was not only efficient, it also can be seen to be the only means by which early societies could flourish. The alternative was either the continuation of wasteful combat, or the relegation of some groups to a continued nomadic life, or both. Thus, cooperation, and its concomitant benefits to participants in agricultural communities, would be Pareto Optimal.

It is widely proposed that the development of agriculture and animal husbandry created the first human experience of surplus.²⁰ This surplus, in turn, accelerated the development of specialization of labor.²¹ Specialization of labor affected the reciprocal entitlements and obligations of three principal groupings: (1) those engaged in agriculture; (2) artisans; and (3) those who undertook domestic and child-rearing obligations. Those engaged in agriculture had, of course, the duty to efficiently and productively produce and to husband the resources and the comestible rewards entrusted to them. Unlike the expectations typical of the hunting and gathering communities, the development of agriculture both permitted and required that what was produced not be consumed immediately, and that when it was consumed, that it not be consumed exclusively by those who produced it. Rather, the expectation for and the duty of those tilling the fields or tending the animals was to harvest the crops and to preserve the harvest, or to slaughter the livestock and to preserve the meat through salting or otherwise, for distribution among the entire community. The artisans were expected to perform such tasks as the creation of the specialized tools that might be associated with chopping, sewing, tilling, the making of clothing, the building of shelter, and more. The artisans' expectation was that, in exchange for their labor, they would partake of the agricultural production of the fields.

¹⁸ The Pareto criteria for wealth maximization analysis are summarized in David W Barnes and Lynn A Stout, *The Economic Analysis of Tort Law* (West Publishing, 1992) 11.

¹⁹ Mark Seidenfeld, *Microeconomic Predicates to Law and Economics* (Anderson Publishing, 1996) 49. For a general description of Pareto optimality principles, see Robin Paul Malloy, *Law and Economics: A Comparative Approach to Theory and Practice* (West Publishing, 1990).

²⁰ Agriculture and animal husbandry will be referred to collectively as 'agriculture'.

²¹ Roberts, above n 15, 51.

The homemakers also might not participate directly in agricultural production, or if they did, they might do so to a lesser extent than those to whom that task would fall principally. The homemakers' primary tasks would include the bearing, raising, and nurturance of children, and the maintenance of a habitable home site, thus freeing both the laborers in the field and the artisans to pursue their work unimpeded of at least the most time consuming obligations of home and child. In return for these responsibilities, the homemakers would rely on the sowers and the reapers, and also the artisans, to share in an equivalence what they had produced.

The significance of these simple group structures, duties, and expectations lay in their promise of and similarity to the more complex duties and expectations that would develop as agriculture permitted the development of larger and more concentrated communities. These larger social or societal groupings would, with the advent of writing and symbolic communication, become the earliest examples of what is now called civilization. And it is in the writings of these early societies that we find articulations of civil responsibility for wrongdoing, or tort law.

III DEVELOPING EXAMPLES OF EFFICIENT FORM AND FUNCTION

A Mesopotamian Law

The watershed discovery and translation of approximately three thousand years of law from the cradle of civilization, framed by the Tigris and the Euphrates Rivers, permitted research, evaluation, and legal synthesis of myriad legal matters. Mesopotamian ancients were, many claim, the first to reduce customary law to written form in an organized and lasting manner.²² The most influential of these laws were collected in the Laws of Hammurabi,²³ the Laws of Ur-Nammu, and the Laws of Lipit-Ishtar.²⁴ The epoch contemplated by these principal bodies of law is approximately 4600 BC to 1600 BC. Although these legal codes were promulgated, published, and republished under the aegises of different rulers and literally millennia, scholars suggest that the 'similarities'²⁵ in the form of the 'academic tradition', the provisions themselves, 'suggest enduring commonalities in the customary law of Babylonia'.²⁶

²² Russ Versteeg, *Early Mesopotamian Law* (Carolina Academic Press, 2000) 3. Several of the references to the principal Mesopotamian codes derive from this work, as well as the translation by L W King: L W King, *The Code of Hammurabi* (Kessinger Publishing, 2004) <<http://eawc.evansville.edu/anthology/Hammurabi.htm>> and C H W Johns, 'The Code of Hammurabi' in *Babylonian and Assyrian Laws, Contracts and Letters* (Kessinger Publishing, 2004).

²³ Versteeg, above n 22. Hammurabi intended that his law reconcile wrongs and bring justice to those aggrieved, with the goal of furthering the economic stability and enhancement of his people.

²⁴ Ibid.

²⁵ Ibid 5, quoting Raymond Westbrook, 'Slave and Master in Ancient Near Eastern Law' (1995) 70 *Chicago-Kent Law Review* 1631, 1634.

²⁶ Versteeg, above n 22, 5, quoting Robert C Eillickson, Charles D A Thorland, 'Ancient Land Use Law: Mesopotamia, Egypt, Israel' (1995) 71 *Chicago-Kent Law Review* 321, 331 (citations omitted).

For present purposes, the legal themes and systems to be discussed will be those of such form and substance as the ancients devoted to systems of customary, normative, and eventually statutory law governing the rights of individuals to be free from wrongful injury, property damage, or coerced takings initiated by others.

Before the Laws of Hammurabi, there were published the Laws of King Ur-Nami (2112–2095 BC). In all of the Mesopotamian law collections, the provisions characteristically begin with an ‘if’ clause (the *prostasis*), and end with a ‘then’ clause (the *apodasis*). Thus, the *prostasis* identifies a circumstance or activity that the lawmakers concluded needed a legal rule, whereas the *apodasis* describes the legal consequences for the creation of such a circumstance or the engagement in such activity.²⁷ This approach bears significant markings of code-based law throughout the ages and is widely followed today.

Review by scholars has revealed numerous examples of remedies for civil wrongs in which Mesopotamian law responded to the *delict* by penalizing, by money judgment, the wrongful disposition (or eradication) of another’s right or vested expectancy. This approach was of particular economic significance in instances where the wronged individual was in a weaker social or economic position than the wrongdoer. Thus, the Laws of Ur-Nami provided that a father whose daughter was promised to a man, but who gave the daughter in marriage to another, must compensate the man (to whom the promise of marriage was made) twice the property value of the property he had brought into the household.²⁸

Similarly, the law emphasized the protection of person, property, and commerce from forced divestiture of a right or a prerogative. Regarding navigation, a collision between two boats on a body of water having a perceptible upstream and downstream would trigger a presumption of fault on the part of the upstream captain, on the logic — faulty or not — that the upstream captain had a greater opportunity to reduce avoidable accidents than his counterpart, as the former would be traveling at a slower speed.²⁹

A subtle interplay between norms of duty, nuisance and causation is evident in the following rule: neighbors were bound by a rule that served to deter letting one’s unoccupied land elevate a risk of trespass or burglary to the neighboring property. The Laws of Lipit-Ishtar provided that where notice had been given by one neighbor that a second neighbor’s unattended property provided access to the complainant’s property by potential robbers, if a robbery occurred, the inattentive neighbor would be liable for any harm to the complainant’s home or property.³⁰ Particularly harsh legal consequences might be visited on the landowner who failed to contain his irrigation canals, as flooding of the water might ‘result not only in leaving crops and cattle dry and parched in one point, but also widespread floods in another part of the

²⁷ Versteeg, above n 22, 11.

²⁸ Versteeg, above n 22, Ur-Nami 9.

²⁹ G R Driver and John C Mills, *The Babylonian Laws* (Clarendon Press, 1952) §§ 431–2, cited in Versteeg, above n 22, 130.

³⁰ King, above n 22, Lipit-Ishtar § 11.

district'.³¹ In the simple case involving only damage to grain, replacement of a like amount might give sufficient remedy. But an unmistakable deterrence of more severe consequences would be clear to those knowing that should the careless farmer be unable to replace the grain, the neighbors might be permitted to sell his property and to sell him into slavery to achieve justice.³²

B Early Religion — The Law of the Torah

It is accepted that much of modern society was suckled at the breast of faith, and that much of mankind's law and morality 'were born of religion'.³³ Often this faith partook of earlier myth, and transformed it to suit the extant needs of the time and the place. And, invariably, the adopted faith adopted strictures against conduct that was inconsistent with the bountiful sustenance of the whole.

The Law of the Torah, with its accompanying interpretation in the Talmud, cannot be described as either ancient or modern, as it is both.³⁴ It represents the longest continuum of international private law that exists. The domain of the Law of the Torah is, strictly speaking, the population of observing Jews. It is, though, of a piece with the same Mosaic law that is the foundation of Christianity,³⁵ and thus its influence has always reached and continues to reach populations and cultures greatly exceeding in number its Jewish adherents.

Israel, and its law, did not differentiate 'between the secular and religious realms'. Rather, all of Jewish life 'was to be lived under Yahweh's command, within his covenant'.³⁶ Included among the contributions of Hebraic law to Western legal development was the recognition that man-made law must give way to God-given, moral law, should the two be in conflict.³⁷ The Torah and its interpretations guide Jews in a very broad spectrum of individual and common pursuits. Naturally, this article is devoted only to such strictures as pertain to the identification of civil wrongs to

³¹ Driver and Mills, above n 29, 50, cited in Versteeg, above n 22, 136.

³² Versteeg, above n 22, Hammurabi § 54. See also Westbrook, above n 25, 1644.

³³ Durkheim, above n 7, 87.

³⁴ Fittingly, religious law — including but not limited to the Law of the Torah — continues to this day to be a part of the weave of both customary law and of national legislation. For example, H W Tambiah states: 'Religion is a source of law through custom or legislation. Difficult questions arise as to the relations between general law and special customary law'. H W Tambiah, *Principles of Ceylon Law* (H W Cave, 1972) 111.

³⁵ The gravitational interplay between Hebrew scripture and Greek philosophy is well-treated in other works. For example, Bertrand Russell, *A History of Western Philosophy* (Simon & Schuster, 1945) 326–7.

³⁶ Bernhard W Anderson, *Understanding the Old Testament* (Prentice-Hall, 2nd ed, 1966) 96. See also Lloyd, above n 11, 49–50, explaining that Hebrew law, revealed law of the Almighty God and embodied in the Law of Moses and later prophets, 'showed that merely man-made laws could not stand or possess any validity whatever in the face of divine laws which the rulers themselves were not competent to reveal or interpret'.

³⁷ Lloyd, above n 11, 50.

others; the remedies for such wrongs; and the sensitivity of such written or traditional law to norms, often distinct in form but typically similar in guidance, of corrective justice and economic efficiency.

The Torah includes the word of God as revealed in the Books of Genesis, Exodus, Leviticus, Numbers, and Deuteronomy.³⁸ These writings, the socio-legal bedrock of Judaism, contain copious treatments, sometimes systematized, of how society ought respond to civil wrongs, and the reasons therefore. Whereas much Western law, particularly modern Western law, is phrased in prohibitory terms, Halakhic law is more apt to treat its society of believers in terms of duty, or put otherwise, 'The observant Jew should ...'.³⁹ Many of these duties are remarkably fuller and more demanding than those recognized in other systematized bodies of law. For example, within the Torah, Leviticus states that a person who stands by while another is put at risk commits a 'crime of omission'.⁴⁰ In the United States and the majority of other legal systems, there is no *ab initio* duty to come to another's aid; rather, such a duty arises only in particular circumstances. The approach stated in Leviticus doubtless describes the higher and more moral road. But might its rationale also resonate in some other social premium important to Jewish society? Apart from obedience to God, another central and seemingly perpetual goal of Jews has been mere survival. It requires no particular boldness to recognize that violence to the persons or the property of members of the Jewish community has always been the subject of closely-held awareness in Jewish communities.⁴¹

A predicate to the advancement of the welfare, progress, and justice of a social group or a state is of course that the group survive as a human community. As the chosen people with limited property of their own, it is proven that the historical Jews were set on by army after army, and it is quite certain that what behavior, from simply cruel to savage, that was not visited on them collectively was surely inflicted on them in discrete, individual and unrecorded incidents. An interpretation that the Law of God required spontaneous protection of other Jews from danger might be seen as a simple and justifiable requirement of the survival of Judaism and its believers.

The Code of the Covenant, set out at Exodus 24:3–8, describes rights and restrictions regarding 'slaves, cattle, fields, vineyards and houses'.⁴² The civil code-like

³⁸ This corresponds to what Christians would later recognize as the first five and similarly named Books of their First Covenant.

³⁹ J David Bleich, *Contemporary Halakhic Problems* (KTAV Publishing House, 1977) 204, n 15, cited in Daniel Pollack (ed), *Contrasts in American and Jewish Law* (KTAV Publishing House, 2001) 226.

⁴⁰ Daniel Pollack et al, 'The Capacity of a Mentally Retarded Person to Consent: An American and a Jewish Legal Perspective' in Daniel Pollack (ed), *Contrasts in American and Jewish Law* (KTAV Publishing House, 2001) 226.

⁴¹ Compare Ernest J Weinrib, 'The Case for a Duty to Rescue' (1980) 90 *Yale Law Journal* 247.

⁴² Roland De Vaux, *Ancient Israel: its Life and Institutions* (John McHugh trans, McGraw-Hill, 1961) 143, cited in Smith and Weisstub, above n 13, 197.

provisions therein are replete with guidance to the community regarding permissible and impermissible community conduct as it affects land, material, and economic transactions. One borrowing another's cloak must return it by nightfall.⁴³ Should one's bull gore a man, the bull is to be stoned.⁴⁴ Even an unworthy thought process that might lead to wasteful bickering or more is enjoined in the admonition 'thou shalt not covet thy neighbor's house ... nor his ass'.⁴⁵

The Talmud and harmonious rabbinical writings are explicit in the condemnation of waste. The 'waste of the resources of this universe is prohibited because of *bal tashit*'.⁴⁶ Such prohibitions include the wasting of food or fuel, the burning of furniture, and the unnecessary killing of animals.⁴⁷

IV EARLY PHILOSOPHY OF IDEAL INDIVIDUAL AND COLLECTIVE PURSUITS

A Hellenic

For a philosophical epoch of greater significance than any other, the Hellenists defined virtue, morality, and ethics in terms that remain the foundation of Western philosophy. Putting aside only a few proponents of distracting philosophic anomalies, the Greek philosophers first identified an ideal of individual behaviors that accented study, modesty in thought and deed, and respect of law. Second, the Hellenist thinkers envisioned a society (at that point a city state) of harmony, accepted strata of skill and task, and, naturally again, respect of law.

However utopian may have been the imagination of such a city state as being led by a politically detached, supremely wise Philosopher-King, the more important instruction is that the Hellenist image of a society and its individual participants was one of social harmony, rewards in the measure of neither more nor less than one's just deserts, and subordination to law. Although undemocratic in many respects including slave-holding, for a pre-democratic, progressive and just ideal evaluated in recognition of its time, the Greece of this era measures up respectably.

Hints of the political circumstances in which Stoics found themselves can be found in the graphics handed down to us from antiquity that portray the various philosophers either speaking to small groups or, from all that appears, to no one at all. There are no representations of them speaking in political groups, or advising political representatives. The reason for this seeming isolation of the philosophers from the political process

⁴³ Exodus 22:25.

⁴⁴ Exodus 18:28.

⁴⁵ Exodus 20:17.

⁴⁶ Daniel Pollack et al, 'Liability for Environmental Damage: An American and a Jewish Legal Perspective' in Daniel Pollack (ed), *Contrasts in American and Jewish Law* (KTAV Publishing House, 2001) 110.

⁴⁷ Shabbat 67b; 129a; Chullin 7b; Sanhedrin 100b at id.

is that by the time of much of the enduring work of the most influential Greek philosophers, political power in the Greek mainland had passed over to the Macedonians. This political powerlessness necessarily affected the focus of many of the philosophers from the politically tinted ‘how can men create a good state?’ to such generally moral issues such as ‘individual virtue and salvation’ and the attendant question ‘how can men be virtuous in a wicked world, or happy in a world of suffering?’⁴⁸

The end sought by Socrates was happiness, which invites the question, a theme of this article, ‘how can a philosophy grounded in the pursuit of “happiness” influence its adherents, much less any larger population, in the ways of moral and efficient civil justice?’ The answer is that to Socrates and other mainstream Hellenic thinkers, happiness could only be achieved through pursuit of the virtuous life, and both the vision and the reality of the virtuous life are suffused with themes not only of morality and justice but also of waste avoidance, and deterrence of unjust enrichment. Socrates’ ethics are permeated by the principle not only of avoidance of doing harm,⁴⁹ but also that the identifying marker of all acts of ‘justice’ was simply ‘returning what was owed’.⁵⁰ For the individual, justice pertained not to the ‘outward man’ but, rather, to the ‘inward man’. The just man ‘sets in order his own inner life, and is his own master and his own law, and [is] at peace with himself’. For the just man, reason governs ‘spirit’ and ‘desire’.⁵¹

To Socrates, self-knowledge was the very essence of virtue. Without such self-knowledge, any man’s accumulation of wealth or power would leave one ‘baffled ... disappointed ... and unable to profit’ from any success. Rejecting the Sophists’ lax attitudes toward generalizable moral or ethical standards, Socrates thought that to be effective, self-knowledge must become so familiar to the adherent that it, and its attendant guidance in virtuous and ethical matters, would be worn like one’s very skin.⁵² To Socrates, wisdom, or self-knowledge, was to be found, at least in one’s early years, through the teaching of wise men. And according to Socrates’ account, there was a broad-based societal subscription to this goal. As all men ‘have a mutual interest in the justice and virtue of one another’, Plato records ‘this is the reason why every one is so ready to teach justice and the laws’.⁵³

To Socrates, temperance conveyed a meaning different from the modern implication of simple forbearance, be it avoidance of alcohol or any other inebriant. Instead, temperance meant the avoidance of ‘folly’ or acting ‘foolishly’. He nevertheless wonders whether virtue was the sum of the parts ‘justice’, ‘temperance’ and ‘holiness’ when he spoke in these words to Protagoras:

⁴⁸ Plato, ‘The Republic: Book IV’, in Scott Buchanan (ed), *The Portable Plato* (Penguin Books, 1977) 448–9.

⁴⁹ *Ibid* 164.

⁵⁰ *Ibid* 159–60.

⁵¹ Russell, above n 36, 230.

⁵² Newton P Stallknecht and Robert S Brumbaugh, *The Spirit of Western Philosophy* (Longmans Green and Co, 1950) 53, 54.

⁵³ Plato, ‘Protagoras’ in Scott Buchanan (ed), *The Portable Plato* (Penguin Books, 1977) 69.

[W]hether virtue is one whole, of which justice and temperance and holiness are parts; or whether all these are only names of one and the same thing: that is the doubt which still lingers in my mind.⁵⁴

Further to the question of why a man should choose the path of justice over injustice, Socrates termed the tension as one of ‘comparative advantage’. He posed the issue as this:

Which is the more profitable, to be just and to act justly and practice virtue whether seen or unseen by gods and men, or to be unjust and act unjustly, if only unpunished and unreformed?⁵⁵

Socrates imagined the ‘tyrannical’ man, one in whom ‘the reasoning ... power is asleep’, and asked ‘how does he live, in happiness or in misery?’ Here Socrates imagines a man of pure impulsivity, a man capable of any ‘folly or crime’. He follows the sad and desperate path of this man, and states that his ‘drunken, lustful, [and] passionate’ habits will require ‘feasts and carousals and revelings’ to satisfy him. Soon such revenues as he may have are spent. In order to continue to feed his uncontrolled desires, the tyrannical man seeks to ‘discover whom he can defraud of his money, in order that he may gratify [his desires]’.⁵⁶ If his parents do not voluntarily submit to his demands, he will try ‘to cheat and deceive them’, and if this fails, he will ‘use force and plunder them’.⁵⁷

The intemperate and unjust man is doomed to a spiral of ever-worsening degradation, Socrates warns. This tyrannical man, Socrates and Adeimantus conclude, is ‘ill governed in his own person’⁵⁸ and knows no true friends, as when they have ‘gained their point’ from another ‘they know them no more’; he never knows ‘true freedom’, as he is a simple instrument of his desires, and is ‘the most miserable’ of men.⁵⁹

Socrates’ encomium of temperance in all pursuits is of course quite analogous to the recognition in later tort theory of the central role of self-restraint. Socrates characterizes as ‘invalids’ those who ‘hav[e] no self-restraint, [and] will not leave of their habits of intemperance’. In essence, Socrates thought temperance could be achieved by ‘a man being his own master’, which is to say, ‘the ordering or controlling of certain pleasures or desires’, and the avoidance of ‘the meaner desires’.⁶⁰

⁵⁴ Ibid, 72, 75–6.

⁵⁵ Plato, above n 48, 451.

⁵⁶ Plato, ‘The Republic: Book IX’ in Scott Buchanan (ed), *The Portable Plato* (Penguin Books, 1977) 625, 628.

⁵⁷ Plato, ‘The Republic: Book X’ in Scott Buchanan (ed), *The Portable Plato* (Penguin Books, 1977) 680.

⁵⁸ Plato, above n 56, 637.

⁵⁹ Ibid 631, 632.

⁶⁰ Ibid.

Socrates compares evil to bodily illness. As a bodily illness can corrupt and destroy bodily health, so, too, can evil destroy a man's soul: 'Does the injustice or other evil which exists in the soul waste and consume her?',⁶¹ and do they not 'by attaching to the soul and inhering in her at last bring her to her death, and so separate her from the body?'⁶² He also subscribes fully to the existence of a heaven and a hell, as is illustrated by the story he tells Glaucon of Er, the son of Armenius, whose body, after he has fallen in battle, is seemingly uncorrupted by death. On the twelfth day, and prior to his burial, he awakens and tells a tale of men being summoned to justice in a mysterious place in which men's deeds are 'fastened on their backs'. The good and the just are led to a 'meadow, where they encamped as at a festival', whereas those found unjust or evil are thrown into a hell in which their punishments are tenfold the average of a man's years, or ten times one thousand in the mythical account.⁶³

Plato's Socrates 'argued for the identity of law and morality'.⁶⁴ Reverence for the law followed from recognition of an implied agreement, to Dennis Lloyd, 'an early form of social contract', for adhering to the law irrespective of the consequences.⁶⁵ Morality, by contrast, would never override the articulated law of the State. While morality might persuade the individual to conclude that the existing law was immoral or unjust, when the two were in conflict, the disputant's 'duty' is 'confined to trying to persuade the state of its moral error'.⁶⁶ In the Hellenic dialogues of Socrates, it is evident that justice entails literally 'called into account' the transgressor, or a pre-Aristotelian expression of corrective justice. As the Sophist Protagoras suggests in Plato's *Protagoras*, the City stands in the shoes of the schoolmaster in giving to 'young men' the laws to be followed. '[T]he laws', states Protagoras, 'which were the invention of good lawgivers living in the olden time; these were given to the young man in order to guide him in his conduct whether he is commanding or obeying'. '[H]e who transgresses them', Protagoras continues, 'is to be corrected, or in other words, called into account'.⁶⁷

Socrates himself speaks even more forcefully of the importance of correcting the defect of misbehavior, and of the deterrent value of punishment. In Book XI of Plato's *Republic*, Socrates tells Glaucon no man 'profits' from 'undetected and unpunished' wrongdoing, as such a man "only gets worse". To Socrates, it is better that the man be detected and punished in order that 'the brutal part of his nature [be] silenced and humanized', and that 'the gentler element in him is liberated'. The man's 'whole soul is perfected and ennobled by the acquirement of justice and temperance and wisdom'.⁶⁸ Socrates discouraged in the most direct terms individual

⁶¹ Plato, above n 57, 680.

⁶² Ibid.

⁶³ Plato, above n 57, 687, 688.

⁶⁴ Lloyd, above n 11 (emphasis added).

⁶⁵ Ibid.

⁶⁶ Ibid.

⁶⁷ Plato, above n 48, 432.

⁶⁸ Plato, above n 56, 655.

miserliness, hoarding, and a spirit of contention and ungoverned ambition. To him, these unworthy characteristics in men were ‘due to the prevalence of the passionate or spirited element’, uncontained by temperance and reason.⁶⁹

To Socrates, the ideal ‘State’ was largely an extrapolation of the ideal man. The State should, Socrates states, have ‘political virtues’ of ‘wisdom, temperance, [and] courage’ that could stand on an equivalence with Socrates’ ideal for the individual.⁷⁰ Identification and description of the fourth virtue, ‘justice’, was more rarified and elusive, and Socrates comments tellingly: ‘[t]he last of those qualities which make a state virtuous must be justice, if only we knew what that was’.⁷¹ A life of virtue and ethics could only be sustained in ‘a law abiding and orderly society’.⁷² Whatever such state-sanctioned justice might be, Socrates commended abidance with existing law, a commitment that ultimately led to his rejection of opportunities to flee his death sentence.

Hellenist thinking could not be reduced to the aphorism ‘virtue is its own reward’. Rather, there were specific rewards associated with a life of virtue, as well as real or imagined disincentives to the adoption of a baser life and the collateral degrading pursuits associated therewith. Time and time again the philosophers stated that a life of excess, be it eating, drinking, or both, incapacitated the actor from realization of the contributions available to and expected of citizens of virtue.⁷³

To both Plato and Socrates, the just man would be content, if not happy, and the unjust man miserable.⁷⁴ In addition, and more specifically, such excesses as invited physical illness and impairment represented a certain departure from God’s, or a god’s, charge to mankind.

For those who might be tempted to depart from a good life, Hellenic writing portrayed strong deterrents. At an individual level, the writings repeatedly allude to the dissipating results of a life of excess, to wit, personal physical deterioration, coupled with personal and communal moral degradation. At such time as man should shed his mortal coil, Socrates and other believers in reincarnation wrote of another reason why a man should choose the path of good. Incapable of disproof and widely

⁶⁹ Plato, ‘The Republic: Book VIII’ in Scott Buchanan (ed), *The Portable Plato* (Penguin Books, 1977) 592.

⁷⁰ Plato, above n 48, 422, 425, 430–1, 434.

⁷¹ Ibid 434.

⁷² Ibid 59.

⁷³ Plato, *Protagoras* (C C W Taylor trans, Oxford University Press, rev ed 1975) 46:
To Protagoras, Socrates spoke of the physical dangers of excess: ‘Don’t you maintain that ... in some circumstances ... when you are conquered by the pleasures of food and drink and sex, you do things though you know them to be wrong?’ ‘Yes’. ... ‘Do you suppose, Protagoras, that they would give any other answer than that they are bad not because they produce immediate pleasure, but because of what comes later, diseases and the like?’ ... And surely in causing diseases they cause pains, and in causing poverty they cause pains.

⁷⁴ Gottlieb, above n 9, 174.

believed, Socrates and others believed that they had lived before in other forms, and that after their demise they would be reincarnated in some animal form.⁷⁵ If a person had led a virtuous life, he would be reincarnated in the form of an animal respected by man, such as a horse. For the man who strayed, his just deserts might well be reincarnation as an insect, perhaps even a dung beetle.

In Plato's version of Socrates, that 'justice' is not just moral but also efficient is quite clear. Even more specifically, the lasting philosophy of this age stated that individual good and justice were, in fact, more than efficient — they were profitable. Socrates states just this: '[o]n what ground, then, can we say that it is profitable for a man to be unjust or self-indulgent or to do any disgraceful act which will make him a worse man, though he can gain money and power?' Happiness and profit inure to the man who, alternatively, 'tame[s] the brute' within, and is 'not be carried away by the vulgar notion of happiness being heaping up an unbounded store', but instead follows the rule of wisdom and law encouraging 'support to every member of the community, and also of the government of children'.⁷⁶

A principal means to the end of justice, to Plato, was education to such a level of legal sophistication that the individual would learn understanding of and respect for the legal process, including such legal process as might pertain to the redress of injury. This is revealed in Socrates' dialogue with Adeimantus in Book IV of *The Republic*. Here Socrates states plainly that it is through education that the individual learns

about the business of the agora, and the ordinary dealings between man and man, or again about dealings with artisans; about insult and injury, or the commencement of actions, and the appointment of juries[.]⁷⁷

Returning what was owed, in effect giving up the actual or conceptual unjust enrichment associated with a wrongful taking, is of course a core model in modern tort for remediation of unconsented-to harm, a concept that is the darling of corrective justice and efficiency advocates alike. It is also part and parcel to the analysis of Aristotle, in *Nicomachean Ethics* Book V, Ch 2, in which 'The Thinker' is credited with laying the corner-stone of the corrective justice principles of today's common law,⁷⁸ although as suggested, the logic has equivalent bearing upon economic

⁷⁵ See discussion of the choices for their future reincarnation by a panoply of demigods by Socrates in his recitation of the story of Er: Plato, above n 57, 694–9.

⁷⁶ Plato, *The Republic of Plato* (Francis McDonald Cornford trans, Oxford University Press, 1969) 318–20.

⁷⁷ Plato, above n 48, 421–2.

⁷⁸ [T]he law ... treats the parties as equal, and asks only if one is the author and the other the victim of injustice or if the one inflicted and the other has sustained an injury. Injustice in this sense is unfair or unequal, and the endeavor of the judge is to equalize it.

Aristotle, *Nicomachean Ethics* (J Welldon trans, Prometheus Books, 1987) 154, discussed in David G Owen, 'The Moral Foundations of Punitive Damages' (1989) 40 *Alabama Law Review* 705, 707–8, n 6.

considerations. Aristotle's understanding was that corrective justice would enable restoration to the victim of the status quo ante, insofar as a monetary award or an injunction can do so.⁷⁹ Under the Aristotelian principle of *diorthotikos*, or 'making straight', 'at the remedy phase the court will attempt to equalize things by means of the penalty, taking away from the gain of the wrong-doer'. Whether the wrong-doer's gain is monetary, or measured in property, or the community's valuation of a personal physical injury consequent to the defendant's wrongful act, by imposing a remedy approximating the actor's wrongful appropriation and 'loss' to the sufferer, 'the judge restores equality'.⁸⁰

Aristotle classified among the diverse 'involuntary' transactions that would invite rectification of 'clandestine' wrongs 'theft, adultery, poisoning ... false witness' and 'violent' wrongs, including 'assault, imprisonment ... robbery with violence ... abuse, [and] insult'.⁸¹ He proceeds to distinguish between excusable harm and harm for which rectification may appropriately be sought. For an involuntary harm, such as when 'A takes B's hand and therewith strikes C', or for acts pursuant to 'ignorance', a more nuanced legal response is indicated. Even for such involuntary acts as 'violat[e] proportion or equality', Aristotle suggests opaquely, some should be excused, whereas others should not be excused. As to voluntary and harmful acts attributable to ignorance, Aristotle distinguishes between acts in which the ignorance is excusable and acts in which the ignorance is not.⁸² The former, which we might today characterize as innocent, would not prompt remediation, whereas the latter would. Thus, Aristotle describes an act from which injury results 'contrary to reasonable expectation' as a 'misadventure', and forgivable at law.⁸³ To Aristotle, an unintentional act⁸⁴ that causes harm, but in which such harm 'is not contrary to reasonable expectation' constitutes not a misadventure but a 'mistake'. To Aristotle, 'mistake' is a fault-based designation. The example used is redolent of the sensibility of what would be termed 'negligence' in today's nomenclature, eg a man throwing an object 'not with intent to wound but only to prick'. This man, although not acting with an intent to wound another in any significant way, would nonetheless be subject to an obligation in indemnity, for to Aristotle, when 'a man makes a mistake ... the fault originates in him'.⁸⁵

⁷⁹ Aristotle, above n 78, 407: 'Therefore the just is intermediate between a sort of gain and a sort of loss, viz, those which are involuntary; it consists in having an equal amount before and after the transaction'.

⁸⁰ Aristotle, *The Complete Works of Aristotle* (Princeton University Press, the revised Oxford translation, 1984) 786.

⁸¹ Aristotle, *Nicomachean Ethics* (book 5, chapter 2), cited in Richard McKeon (ed), *Introduction to Aristotle* (University of Chicago Press, 1947) 402.

⁸² *Ibid* 414.

⁸³ *Ibid* 415.

⁸⁴ An act that 'does not imply vice': *ibid*.

⁸⁵ *Ibid*.

Aristotle's famous 'Golden Mean' hypothesizes that virtue analyzed linearly is the mean between two extremes. Either extreme is a vice. So, for example, if appropriate self-sustenance is a virtue, then it follows that, at one extreme, self-denial to the point of ill health is a vice. At the other extremity, gluttony is a vice. Importantly, Aristotle does not propose distributive justice, in the sense that a man may remedy his antecedent unequal position vis-à-vis another.⁸⁶

B Roman Law

It is recognized generally that the Romans added little to the metaphysics of law. Nevertheless, Roman law represents the watershed between the law of ancient society and that of modern society. As suggested by Sir Henry Sumner Maine, the rights and the duties under law of ancient society derived from status, or 'a man's position in the family', whereas under Roman law and thereafter, rights and duties 'derived from bilateral arrangements'.⁸⁷

Regarding *delicts*, or harms that were neither crimes nor grounded in contract, it became the special province of Roman lawyers and lawmakers to record and categorize a sprawling array of specific wrongs and consequent remedies. This approach of Roman law would become the origin of code-based law that governs European and Latin American lawmaking to this day.

Cicero, the Roman orator, wrote of an ethic that sounded simultaneously in terms of corrective justice and efficiency-deterrence. In *On Moral Duties* he wrote that even after 'retribution and punishment' have been dealt to the transgressor, the person who has been dealt the wrong owes a duty to bring a close to any such misadventure by permitting a gesture such as repentance or apology. From the extension to the wrongdoer of the opportunity to apologize or to repent could be reaped the immediate good of reducing the likelihood that he would 'repea[t] the offense', as well as the broader and eventual good of 'deter[ing] others from injustice'.⁸⁸

Cicero further propounded a cluster of maxims that, if followed, could conduce to Pareto Superior changes, in the sense that the actor would be no worse off and the affected party would be better off. As to persons beyond a benefactor's core family or kinship group, to Cicero there existed a duty to the entire world as to such things 'we receive with profit and give without loss'. Thus, in order that we may receive such blessings as are identified in the maxims such as 'keep no one from a running

⁸⁶ To Aristotle:

Justice (contrary to our own view) implies that members of the community possess unequal standing. That which ensures justice, whether it is with regard to the distribution of the prizes of life or the adjudication of conflicts, or the regulation of mutual services is good since it is required ... for the continuance of the group. Normativity, then, is inseparable from actuality.

Polanyi, above n 14, 83.

⁸⁷ Polanyi, above n 14, 82–3.

⁸⁸ Cicero, 'On Moral Duties' in Moses Hadas (ed), *The Basic Works of Cicero* (Modern Library, 1951) 16.

stream' or 'let anyone who pleases take a light from your fire' or 'give honest advice to a man in doubt', Cicero writes, it follows that we must be willing to give likewise of the same in order to 'contribute to the common weal'.⁸⁹

Under Roman law, two of the *delicts* of greatest importance were damage to property, real and personal, and personal physical harm to others, giving rise to the *action injuriarum*. The victim could bring an action for 'profitable amends', or money damages, or 'honorable amends', which is to say, a formal and public apology. The latter remedy would most likely arise in a setting of a dignitary tort, such as defamation. Roman jurists and the Roman legal community were committed to the identification of the delineation between what is 'just and what unjust', and therefore the Institutes of Justinian and other sources of Roman law reflected an endeavor to 'give each man his due right', and comprise 'precepts' to all Romans 'to live justly, not to injure another and to render to each his own'.⁹⁰

The Institutes included numerous strictures against imposing one's will over the rights of a neighbor, and strong deterrents for the disregard thereof. As to urban estates, in Book III, Title II Para 2 there was a prohibition on the obstruction of a neighbor's view,⁹¹ a rule bearing a resemblance to a former English and American rule that limited a neighbor's liberty to interfere with 'ancient lights'.⁹² In another notable example, pertaining to what would today be called the law of private nuisance or trespass, Roman law detailed a preference that adjoining landowners bargain in advance for agreement as to contemporaneous uses of land that might trigger a dispute. In Book III Para 4, the Institutes provide that one 'wishing to create' such a right of usage 'should do so by pacts and stipulations'. Further, a testator of land might impose such agreements on his heirs, including limitations on building height, obstruction of light, or introduction of a beam into a common wall, or the construction of a catch for a cistern, an easement of passage, or a right of way to water.⁹³

These last two examples reflect an encouragement of *ex ante* bargaining over economically wasteful *ex post* dispute resolution. Additionally, permitting the testator to bind his heirs to any such agreement was economically efficient in a manner akin to the approach taken later and famously by Justice Bergen in the cement plant nuisance case of *Boomer v Atlantic Cement Co.*⁹⁴ In *Boomer*, discussed below, the court's award of damages ensured that there would be a one-time resolution of the dispute by requiring that the disposition of the claim be entered and recorded as a permanent servitude on the land.

⁸⁹ Ibid 23.

⁹⁰ J A C Thomas (ed), *The Institutes of Justinian: Text, Translation and Commentary* (North-Holland Publishing, 1975) Book I, Preamble, [1], [3], 84, 85.

⁹¹ Ibid, The Digest (or Pandects) Book III Title II [2], [3].

⁹² See, eg, *Robeson v Pittinger*, 2 NJ Eq 57 (NJ Eq 1838). The doctrine is no longer followed generally in American law.

⁹³ Ibid, Book III [4], 84, 85.

⁹⁴ 26 NY 2d 219 (NY, 1970) ('*Boomer*').

VI MODERN ASSIGNMENT OF ECONOMIC NORMS

A Customary Law

The organized law of the modern state is a fairly recent phenomenon when compared to the existence of effective and nuanced customary law around which pre-modern societies organized. For ancient and modern societies alike, and be the law written or unwritten, law represents a cohering of the ‘underlying social norms which determine much of its functioning’. This customary law has been described as ‘living law’.⁹⁵

In the history of tort law, customary law has from time before time given effect to group norms and other socio-legal principles.⁹⁶ At such later time as a culture or a nation-state has begun to render its law in the form of written adjudicatory rulings, or legal codes, customary law characteristically diminishes in its significance as an engine for resolution of disputes. With notable French influence, the theories of Roman law became ascendant,⁹⁷ and the recitation of and reliance upon customary norms receded proportionately. Yet in many societies, even today, customary law continues to inform legal development. In some settings, customary law sets parameters for later legal development, or even precludes later law that would contradict earlier custom.

The influence of Roman law on the development of European, Latin American, and Anglo-American law is commonly acknowledged. The lasting effects of customary law, including Roman law, and one of its hybrids, Roman-Dutch law, are well-illustrated in the experience of Ceylon (modern Sri Lanka), the customary law of which reveals a systemic commitment to wealth maximization, avoidance of waste, and deterrence of behaviors inconsistent therewith.

For Celanese customary law, and for that matter any customary law, to be considered valid for the purposes of modern adjudication, it must be (1) reasonable; (2) consistent with the law contemporary to that society; (3) universal in application; and (4) grounded in antiquity.⁹⁸ Although the first and second standards might at first glance seem to subordinate customary law into insignificance, an additional look makes this approach appear more sensible. This approach can be justified on two grounds. First, as with common-law adjudication, as no court is required to apply common law that is unreasonable, it would be illogical to require application of customary law that was not reasonable. Secondly, as both common law and customary law claim lineage in a society’s reasoned conclusions as to legal standards best suited to societal wellbeing, customary law that was at war with common law on the same or a similar subject would be presumptively defective in either its rationality or in its claimed representational authenticity.

⁹⁵ Lloyd, above n 11, 227 (citation omitted).

⁹⁶ Ibid, discussing the sources and growth of custom.

⁹⁷ John Richard Green, *A Short History of the English People* (Macmillan, 1874) 204.

⁹⁸ Tambiah, above n 34, 87–9. See generally, Tambyah Nadaraja, *The Legal System of Ceylon in its Historical Setting* (Brill Archive, 1972).

In some legal systems, legal scholars remain oracles of greater or lesser significance. This was true of the Roman-Dutch tradition, in which schools of legal scholarship, or the scholars themselves were influential. There existed two schools of writers: (1) Grotius, van Leeuwen and Voet, who emphasized the Roman law antecedents of the developing hybrid law; and (2) an ‘historical school’ that emphasized custom as the appropriate principal source of the law.⁹⁹ Although the Napoleonic Code superseded the Roman-Dutch law in Holland itself in the early nineteenth century, the great Dutch East and West Indian Trading Companies carried Roman-Dutch law into their settlements. So strong was the influence of custom in the Roman-Dutch tradition that, in principle at least, a statute could be rendered nugatory or obsolete by sufficient proof of a conflicting custom.¹⁰⁰

Under the so-called *Lex Aquilia* of Roman-Dutch law, the Aquilian action required the claimant’s showing of a wrongful act, patrimonial loss, and the defendant’s fault, because of either the intentional nature of the act or negligence. Borrowing from British law applicable to unintentional injuries, the law courts of Ceylon adopted the British concept that the plaintiff’s claim in negligence must include proof that the defendant owed a duty to the plaintiff.¹⁰¹

In Aquilian actions, no compensation would be awarded in the absence of physical injury, with physical injury classically defined as excluding emotional distress or dignitary harm. Regarding injuries caused by the positive act of the defendant, should a person be in possession of a thing, including a chattel or an instrumentality, that had the potential for causing harm if not stewarded with care, the actor, owner, or manager would have a positive duty to exercise such care. Should another be injured because of the failure to take such care, liability could be imposed. Even a mere omission to act might be a stimulus to liability if the actor’s omission was ‘connected with some prior positive act’. Accordingly, a remedy might be available under the *Lex Aquilia* if the defendant had earlier ‘created a potentially dangerous state of things’, and the failure to correct that caused the claimant’s harm.

Various dimensions in the Roman-Dutch tradition recognized the society’s commitment to the integrity of persons and property from forced takings. Assault was an *injuria*, and therefore redressable, on a showing of *contumelia*. For grazing animals that damaged another’s property, if the animal’s transgression involved the ‘animal acting contrary to nature of its class’, the owner might be required to pay damages, or even be confronted with the potentially stronger deterrent of giving up

⁹⁹ Tambiah, above n 34, 87–9; Green, above n 96, 148. The opinions most often referenced can be found in Johannes Voet, *Commentarius Ad Pandectas* (William Stoney trans, Oxford University Press, 1896) [trans of: *Commentarius Ad Pandectas* (first published 1704)]; Hugo Grotius, *Inleiding tot de Hallanesche Rechtsgeleertheyd* (R W Lee trans, Clarendon Press, 1953) and Simon Van Leeuwen, *Simon Van Leeuwen’s Commentaries on Roman-Dutch Law* (J G Kotze trans, Sweet and Maxwell, 1921).

¹⁰⁰ Tambiah, above n 34.

¹⁰¹ *Ibid* 99.

the animal. Surely, too, a strong message of deterrence is found in the rule that a person finding another's animals on that person's property could impound them.¹⁰²

For 'intentional' wrongs, the intentional torts of today, the requisite intent, or *dolus*, was provided by the defendant's desire to accomplish the act, irrespective of whether he was aware that the act constituted an invasion of the plaintiff's rights. '*Culpa*' was interpreted as a 'violation of a duty that [is] imposed by law',¹⁰³ an approach revealing the influence of the English common law requirement that the tort plaintiff prove duty. The respondent could avoid liability by showing that the injury could not have been avoided even by the exercise of reasonable care. Furthermore, in order to avoid unjust enrichment, persons could not recover for claims arising from acts or activities to which they consented voluntarily.

For intentional torts such as false imprisonment, the third requirement of the *Lex Aquila*, that of foreseeability, would be satisfied by showing that the defendant intended the act. Then, as it is today, a reflection of the rigorous economic guardianship of customary law, Dutch-Roman law gave primacy to protection of property and economic rights, and imposed an almost automatic requirement of disgorgement of any unjust enrichment associated with the wrongful interference therewith. Trespass, or the willful and forcible entry into another's property, constituted *injuria*. As has been true for any successful socio-economic unit, the Roman-Dutch tradition recognized the rights of a person to protect his property from any form of unjust interference.¹⁰⁴

At the same time, it was recognized that a landlord owed a duty to his tenants to take reasonable steps to protect them from injury caused by unsafe conditions on the land.¹⁰⁵ In what could be loosely styled as a proscription of public nuisance, the Roman-Dutch customary law removed earlier Praetorian edicts prohibiting certain animals from sharing public places, and replaced the strict prohibitions with rules requiring the payment of damages.¹⁰⁶ Private nuisance, in turn, was seemingly remediable in an action for damages or for equitable relief.¹⁰⁷

Roman-Dutch customary law includes at least one example of the law and its official apparatus not being required to stand idly by to await the social costs of an

¹⁰² Ibid, 392–5, 399, 418, 420. For damage caused by trespassing dogs, the claimant would be required to show scienter. Within this approach there could be seen a strong overlay of moral blameworthiness:

It is doubtful whether Roman and Roman-Dutch writers regarded negligence objectively or subjectively, but partly under the influence of canon law, and its offspring natural law, the modern systems based on Roman law took culpa to imply moral blameworthiness:

Ibid, 397.

¹⁰³ Ibid 397.

¹⁰⁴ Ibid 142, 397, 399, 418.

¹⁰⁵ Ibid 399.

¹⁰⁶ Ibid 422.

¹⁰⁷ Ibid 396.

accident that will occur or that will continue to occur in circumstances in which the parties had not reached a prior agreement as to risks and rewards. Should a neighbor come to fear that a dangerous condition existed in his neighbor's house that, left unabated, might cause damage to the property of the complainant, the complainant could bring an action in what might today be called anticipatory nuisance demanding the neighbor's payment of security against such prospective and potential harm.¹⁰⁸

B Modern United States Accident Law

The analysis of tort law has long emphasized its original and lasting tenets in the logic of corrective justice and morality. Nevertheless, economists, political scientists and legal scholars repair with increasing frequency and interest to the examination of economic truths within the function of injury law, including evaluation of evolving decisional law against the measure of whether such decisions adhere, explicitly or silently, to goals of economic efficiency.

It is established that tort law is devoted to the protection of persons and property from unreasonable risk of harm; and the actor's liability in tort is limited by concepts of reasonable foreseeability. Putting aside the cabined domain of truly strict liability, modern accident law is concerned primarily with the provision of reparations to persons suffering personal injury or property loss because of fault, with fault conventionally defined as a failure of others to act with due care under the circumstances.

Much of United States negligence law in accident cases has ordained a finding of liability in negligence upon the plaintiff establishing duty, breach, legal (or proximate) cause, and damages. While each of the four aspects of the plaintiff's prima facie case are influenced by economic considerations, it is breach that is most suited to the polycentric efficiency considerations of the individual and social benefits, together with the individual and societal burdens, of the actor's conduct (or its cessation).¹⁰⁹

Courts in numerous jurisdictions employ the formulation of Judge Learned Hand, or a harmonious risk-utility model. This primitive but enormously influential calculus

¹⁰⁸ Ibid 422.

¹⁰⁹ From the start, the concept of negligence has been based on the notion of 'reasonableness', predicated on the idea that proper decisions involve selecting the proper balance of expected advantages and disadvantages, of expected benefits and costs. David G Owen, 'Defectiveness Restated: Exploding the "Strict" Products Liability Myth' (1996) *University of Illinois Law Review* 743, 7545. See American Law Institute, *Restatement (Third) of Torts: Products Liability* (1998) reporters' notes to cmt (a). See, eg, *Rodriguez v Spencer*, 902 SW 2d 37, 46 (Tex Ct App, 1995), a negligence suit brought against the parent of a minor who engaged in a fatal beating of a homosexual. The Court factored in the social utility of the parent's conduct in the context of there being no evidence that 'created a genuine issue of fact material to Spencer's duty to prevent her son from engaging in behavior so unforeseeable as the atrocity against Paul Broussard'.

was offered in a negligence context by Hand J in the opinions in *United States v Carroll Towing Co.*¹¹⁰ and *Conway v O'Brien*.¹¹¹ In those two cases, the court stated that 'the degree of care appropriate to a situation is the result of the calculus using three factors: the likelihood that the conduct will injure others, multiplied by the seriousness of the risk if it happens, balanced against the burden of taking precautions against the risk'. The formula is known to many as B (burden) < P (probability of harm) L (magnitude of loss, should it occur). The Learned Hand approach can be conformed to a utilitarian analysis by visualizing B as encompassing not only the particular burden of precautionary measures upon the actor, but also the burden upon society if the conduct must either be eliminated due to liability rules, or made more expensive by requiring precautionary measure and therefore beyond the economic reach of many.

Applying Hand J's negligence heuristic device, a precaution is efficient when its cost is lower than its expected benefit. A party behaves unreasonably if she would have been able to avoid the harm by investing in an efficient precaution and did not, while no more efficient precautions were available to other parties.¹¹² Of Hand J's formula, Posner states plainly: '[t]his is an economic test',¹¹³ a conclusion widely endorsed.¹¹⁴

Judge Posner claims further that most tort law rules further economic efficiency. He does not claim that judges write opinions 'in the language of welfare economics or social cost, or even that judges consciously employ rules to maximize the size of the

¹¹⁰ 159 F 2d 169, 173 (2nd Cir, 1949).

¹¹¹ 111 F 2d 611, 612 (2nd Cir, 1940).

¹¹² Ehud Guttel, 'The (Hidden) Risk of Opportunistic Precautions' (2007) 93 *Virginia Law Review* 1389, 1390.

¹¹³ Richard A Posner, *Economic Analysis of Law* (Little, Brown and Co, 2nd ed, 1977) 122 (citations omitted). He continues:

The burden of precautions is the cost of avoiding the accident. The loss multiplied by the probability of the accident is the expected accident cost, ie, the cost that the precautions would have averted. If a larger cost could have been avoided by incurring a smaller cost, efficiency requires that the smaller cost be incurred.

¹¹⁴ See, eg, Michael D Green, 'Negligence = Economic Efficiency: Doubts >' (1997) 75 *Texas Law Review* 1605, 1612: 'the economic cost of that untaken precaution and the expected accident toll if the precaution is not taken ... must be compared with each other in a risk-benefit test'. See also Bruce Chapman, 'Corporate Tort Liability and the Problem of Overcompliance' (1996) 69 *South Carolina Law Review* 1679, 1690:

in a negligence action the plaintiff is required to show, first, that some untaken precaution would have prevented the injury had it been taken and, second, that it was reasonable to require that such a precaution be taken (for example, that the taking of the precaution would pass ... a Learned Hand test).

See further, Mark F Grady, 'Discontinuities and Information Burdens: A Review of The Economic Structure of Tort Law by William M Landes and Richard A Posner' (1988) 56 *George Washington Law Review* 658, 661, showing that performing risk-utility analysis in tort cases is conventionally understood to require exploring the efficiency of untaken precautions.

society's pie'. Rather, he contends that 'people can apply the principles of economics intuitively — and thus "do" economics without knowing they are doing it'.¹¹⁵

Of particular economic significance is the Hand formulation's explicit reference to the 'social utility' of the injury-producing conduct. Thus conduct of a high social utility is incrementally more likely to satisfy the standard of 'breach' than an activity of lesser significance. Inclusion of 'social utility' among the economic considerations identified here is pregnant with the converse question: what would be the societal burden if the activity were barred or made more expensive by virtue of liability awards? Ambitiously, the Louisiana Supreme Court in *Pepper v Triplet*¹¹⁶ compared the a court's task in a negligence action to that of a lawmaker, charged with task of

taking into account all of the social, moral, economic and other considerations as would a legislator regulating the matter, [an] analysis [that] is virtually identical to the risk-utility balancing test used in both *negligence* and products liability theories.¹¹⁷

Thus, separate mention is properly made of products liability actions in which many courts have concluded that any 'social utility' evaluation must take into account the possibility that the costs of reducing a risk may result in the social costs of pricing a product out of the market altogether. In *Thibault v Sears, Roebuck & Co.*,¹¹⁸ a suit arising from the plaintiff's injury when his foot slipped under a lawn mower carriage during operation, when evaluating social utility and the cost of preventative measures, the New Hampshire Supreme Court made clear that cost was not merely a question of the unit cost of adding a more protective feature, but also the cost to the public at large should the remediation require a higher cost, and the cost to the public should such an elevated cost put the product out of reach of a proportion of the consuming public, rendering the design change economically infeasible.¹¹⁹

In accident cases applying ordinary negligence principles, courts routinely apply a risk-utility standard of interrelated factors tracing to a greater or lesser extent that

¹¹⁵ See generally William M Landes and Richard A Posner, *The Economic Structure of Tort Law* (Harvard University Press, 1987) 23. Today law and economics 'is like the air you breathe. It's just pervasive', according to David Weisbach, a University of Chicago Law professor: Peter Coy, 'Will Success Spoil the Chicago School?', *Bloomberg Business Week*, 7 June 2012 <<http://www.businessweek.com/articles/2012-06-07/will-success-spoil-the-chicago-school>>.

¹¹⁶ 864 So 2d 181, 1923 (La, 2004), citing *Boyer v Seal*, 553 So 2d 827, 834–6 (La, 1989).

¹¹⁷ *Pepper v Triplet* 864 So 2d 181, 192–3 (La, 2004).

¹¹⁸ 395 A 2d 843 (NH, 1978).

¹¹⁹ In determining unreasonable danger, courts should consider factors such as *social utility* and desirability: John W Wade, 'On the Nature of Strict Tort Liability for Products' (1973) 44 *Mississippi Law Journal* 825, 837. At 845:

[t]he utility of the product must be evaluated from the point of view of the public as a whole, because a finding of liability for defective design could result in the removal of an entire product line from the market.

employed in *Gonzalez v O'Brien*,¹²⁰ a negligence action arising from a farm tractor incident. While finding against the plaintiff, the court described a risk-utility test that included consideration of 'the social utility of the O'Brien brothers' conduct [as] high (running a ranch)'.

Commentators by now commonly identify a common law tropism towards efficiency.¹²¹ Importantly, scholars have also concluded that efficient rules of law actually predict not only efficient litigation, but also settlement, strategies,¹²² the common law of rescue, salvage, Good Samaritan assistance,¹²³ and the economic loss rule.¹²⁴

A leading exponent of the efficiency role of the common law of tort has been Guido Calabresi. Calabresi has argued persuasively that in matters of compensation for accidents, civil liability should ordinarily be laid at the door of the 'cheapest cost avoider', the actor who could most easily discover and inexpensively remediate the hazard. Together with A Douglas Melamed, Calabresi has counseled that considerations of economic efficiency dictate placing the cost of accidents 'on the party or activity which can most cheaply avoid them'.¹²⁵ Ordinary economic rationales also have described the role of compensatory damages as an effective means of

¹²⁰ 305 SW 3d 186 (Tex Ct App, 2009). In *Ambrose v McLaney*, 959 So 2d 529, 538 n 11 (La Ct App, 2007), a premises liability suit brought by a tenant who sustained injuries in a fall down an exterior stairway, the court applied risk-utility considerations that included '(1) the utility of the complained-of condition; ... and (4) the nature of the plaintiff's activities in terms of its social utility or whether it is dangerous by nature' (citations omitted).

¹²¹ See, eg, George L Priest, 'The Common Law Process and the Selection of Efficient Rules' (1977) 6 *Journal of Legal Studies* 65; Ramona L Paetzold and Steven L Willborn, 'The Efficiency of the Common Law Reconsidered' (1991) 14 *George Mason University Law Review* 157.

¹²² Wes Parsons collected scholarship revealing in fact the broad range of cost internalization achievements of evolving common-law doctrine, including commendation of the common law of accidents as 'promot[ing] efficient resource allocation'. Wes Parsons, 'The Inefficient Common Law' (1983) 92 *Yale Law Journal* 862.

¹²³ William A Landes and Richard A Posner, 'Salvors, Finders, Good Samaritans and Other Rescuers: An Economic Study of Law and Altruism' (1978) 7 *Journal of Legal Studies* 83, 128.

¹²⁴ William Bishop, 'Economic Loss in Tort' (1982) 2 *Oxford Journal of Legal Studies* 1, 2–3.

¹²⁵ Guido Calabresi and A Douglas Melamed, 'Property Rules, Liability Rules, and Inalienability: One View of the Cathedral' (1972) 85 *Harvard Law Review* 1089, 1096–7. See also Mark C Rahdert, *Covering Accident Costs: Insurance, Liability and Tort Reform* (Temple University Press, 1995) 29, 32–3. Marco Jimenez considers:

Under this approach ... a court need 'merely calculate the costs and the benefits of an activity to decide whether an injurer [is] negligent', and need not be concerned with determining what a virtuous or reasonable person in the defendant's position would have foreseen, or whether a tortfeasor's actions were intrinsically right or wrong. Rather ... a court need only determine which party is the 'cheapest cost avoider' (ie, 'the actor who

discouraging a potential tortfeasor from bypassing the market, and by their substandard or risk-creating conduct, injuring an unconsenting third party. Such conduct is wasteful in terms of identifiable accident costs as it is better, theoretically at least, to pressure the actor into bargaining with any willing and knowing other for the right to expose him or her to risk.¹²⁶

The Ninth Circuit in *Union Oil v Oppen*¹²⁷ adopted Calabresi and Melamed's 'least cost avoider' approach. *Oppen* was a California coastal oil spill case in which the court allowed commercial fishermen to recover from the defendant their business losses caused by lost fishing opportunity during a period of pollution. Applying the 'best or cheapest cost avoider' approach, the Appeals Court followed Calabresi and Melamed's predicate that it 'exclude as potential cost avoiders those group activities which could avoid accident costs only at extremely high expense'.¹²⁸ This approach militated against imposing the cost of prevention or repositioning the loss upon the consumers (fishermen or seafood purchasers) in the form of precautionary measures (whatever they might hypothetically be), or by first party insurance. Placing responsibility for the loss on the defendant oil company, the court explained:

the loss should be borne by the party who can best correct any error in allocation, if such there be, by acquiring the activity to which the party has been made liable. The capacity to 'buy out' the plaintiffs if the burden is too great is, in essence, the real focus of Calabresi's approach. On this basis, there is no contest — the defendant's capacity is superior.¹²⁹

Illustrative too is the Third Circuit's decision in *Whitehead v St Joe Lead*,¹³⁰ a lead poisoning case in which the defendants included the suppliers of lead to plaintiff's industrial employer. Reversing summary judgment, the court concluded that

it may well be that suppliers, acting individually or through their trade associations, are the most efficient cost avoiders. Certainly it could be found to be inefficient for many thousands of lead processors to individually duplicate the

¹²⁵ *cont'd*

could most easily discover and inexpensively remediate the hazard'), and then place the cost of accident prevention on this person to encourage them to take only those precautions that are economically feasible:

Marco Jimenez, 'Finding the Good in Holmes's Bad Man' (2011) 79 *Fordham Law Review* 2069, 2094–5 (citations omitted).

¹²⁶ Today, one cannot help but think of the newest 'trash' television shows and derivatives thereof. Cf Guido Calabresi, 'The Decision for Accidents: An Approach to Nonfault Allocation of Costs' (1965) 78 *Harvard Law Review* 713, 716.

¹²⁷ 501 F 2d 558 (9th Cir, 1974).

¹²⁸ *Ibid* 569.

¹²⁹ *Ibid* 569–70.

¹³⁰ 729 F 2d 238 (3rd Cir, 1984).

industrial hygiene research, design, and printing costs of a smaller number of lead suppliers.¹³¹

Whichever gloss is placed on economic analysis — its deterrent effect, or its ability to reduce accident costs — its concepts can be understood ‘even at the rudimentary level of jurists’, according to Judge Patrick Higgenbotham. In *Louisiana ex rel Guste v M/V Testbank*,¹³² the renowned vessel collision case involving claims for economic loss not accompanied by physical damage to a proprietary interest, the Fifth Circuit Court of Appeals justified its refusal to permit such recovery and continued its adherence to the economic loss doctrine of *Robins Dry Dock & Repair v Flint*.¹³³ The court reasoned that permitting liability for the ‘unknowable’ amounts that might be posed as economic loss claims arising from any substantial mishap would erode the efficient deterrent effect of such a tort rule, as a rational, wealth-maximizing actor would be unable to gauge the optimal precautionary measures for avoidance of a predictable accident cost. In Judge Higgenbotham’s words:

it is suggested that placing all the consequence of its error on the maritime industry will enhance its incentive for safety. While correct, as far as such analysis goes, such *in terrorem* benefits have an optimal level. Presumably, when the cost of an unsafe condition exceeds its utility there is an incentive to change. As the costs of an accident become increasing multiples of its utility, however, there is a point at which greater accident costs lose meaning, and the incentive curve flattens.

¹³¹ Ibid 575. See also *Ogle v Caterpillar Tractor*, 716 P 2d 334, 342 (Wyo, 1986) (‘*Ogle*’). The court stated:

When a defective article enters the stream of commerce and an innocent person is hurt, it is better that the loss fall on the manufacturer, distributor or seller than on the innocent victim ... They are simply in the best position to either insure against the loss or spread the loss among all consumers of the product.

Ogle was later described by the Wyoming Supreme Court as an indication of how strict liability ‘introduced economic analysis to tort law’: *Schneider National v Holland Hitch*, 843 P 2d 561 (Wyo, 1992) (‘lowest cost avoider’). The Court in *Schneider National* proceeded to analogize *Ogle*’s ‘risk allocation’ theory to a ‘cheapest cost avoider’ approach. See also *Wilson v Good Humor*, 757 F 2d 1293, 1306 n 13 (DC Cir, 1985), identifying but not pursuing cheapest cost avoider analysis in an action brought by parents of a child who was fatally injured while crossing a street to meet an ice cream vending truck.

Consistent authority is found in the insurance declaratory judgment context. The dissenting opinion in *Insurance Co of North America v Forty-Eight Insulations*, 633 F 2d 1212 (6th Cir, 1980) proposed a ‘discoverability’ rule for triggering insurance carrier coverage of asbestos claims, asserting that this approach would, relying on a least cost avoider rationale, provide incentives to reduce accident costs. Specifically, the dissent reasoned:

The more ‘early’ insurers that are liable upon a victim’s exposure, the more likely it is that the potential harm will be discovered and the public warned. If an insurer sees that the product poses some risks, he may raise premiums accordingly. This may ultimately cause the manufacturer to remove the product from the market or to give better warnings in order to lower insurance premiums. This in turn reduces accident costs.

¹³² 728 F 2d 748 (5th Cir, 1985) (‘*Testbank*’).

¹³³ 275 US 303 (1927).

When the accident costs are added in large but unknowable amounts the value of the exercise is diminished.¹³⁴

In the setting of environmental harm, notions of corrective justice and utilitarianism (or efficiency and equity¹³⁵) have long coexisted uneasily. Originally, even the most economically powerless landholder could seek and secure an injunction against a neighboring activity that interfered substantially with the plaintiff's use of property. Numerous early decisions evidenced a judicial unwillingness to 'balance' injuries, that is, to weigh the defendant's cost and the community hardship in losing the industry against the often modest provable harm to plaintiff's often ordinarily small and noncommercial property. As the New York Court of Appeals stated in *Whalen v Union Bag & Paper*,¹³⁶ to fail to grant the small landowner an injunction solely because the loss to him, in absolute terms, was less than would be the investment-backed loss to the nuisance-creating business and lost employment within the community, would 'deprive the poor litigant of his little property by giving it to those already rich'.

In contrast, the modern rule governing injunctions, including environmental injunctions, might seem coldly utilitarian. The *Restatement (Second) of Torts* § 936 lists factors for injunction issuance, which expressly include weighing of 'the nature of the interest to be protected', thus presumably inviting an elevation of plaintiff's bona fides in cases in which the court considers the activity meritorious (perhaps a recreational area for Alzheimer's patients) and a devaluation in which the court deems it less valuable (perhaps an automobile scrapyard). Along similar lines, hardship to the defendant of ceasing or changing its activity, and 'the interests of third persons and of the public' are proper considerations.¹³⁷ The reference to potential burdens upon third persons and the public represent a clear invitation to introduce concepts of social costs into environmental damage litigation.

Representative of such an approach is the result reached in *Boomer*,¹³⁸ a decision known to legions of law students. *Boomer* involved a large-scale industrial nuisance in the form of airborne cement dust emanating from an upstate New York cement plant. In the lower court, a nuisance was found, and temporary damages awarded, but the plaintiffs' application for an injunction was denied. Recognizing that to deny the injunction would depart from *Whalen*'s corrective justice/no balancing approach discussed earlier, the court nevertheless adopted a utilitarian approach that weighed the hardships imposed on the plaintiffs against the economic consequences of the requested injunction. In what might be described as a split decision, the court denied

¹³⁴ *Testbank* 728 F 2d 748, 1029 (5th Cir, 1985).

¹³⁵ See A Mitchell Polinsky, 'From Political Philosophy to Game Theory' in A Mitchell Polinsky, *An Introduction to Law and Economics* (Little, Brown and Co, 2nd ed, 1989).

¹³⁶ 208 NY 1 (1913).

¹³⁷ *Restatement (Second) of Torts* § 936(1)(e)–(g).

¹³⁸ 26 NY 2d 219 (1970).

the injunction and awarded permanent, one-time damages that would be recorded as a continuing servitude on the land. The court explained:

The ground for denial of injunction, notwithstanding the finding both that there is a nuisance and that [the] plaintiffs have been damaged substantially, is the large disparity in economic consequences of the nuisance and of the injunction.¹³⁹

Boomer permits examination of modern nuisance law in terms of a cost-benefit or utilitarian rationale that vindicates what the New York court concluded were the overall best economic interests of the community. However, other important elements to an economic analysis of nuisance law are at play, to wit, the elements of social cost. In the introductory paragraph to *The Problem of Social Cost*,¹⁴⁰ Ronald H Coase illustrates the operation of social cost analysis by employing the example of a factory emitting demonstrably *harmful* pollutants — in this sense an important distinction with *Boomer*. Coase suggests that application of pure economic principles might prompt economists to conclude that it might be desirable to have the factory pay damages, proportionate or otherwise, or even to have the factory shut down. Such results, he proposes, may be ‘inappropriate, in that they lead to results which are not necessarily, or even usually, desirable’.¹⁴¹

It is received wisdom that in order to be efficient, any tort rule associated with a phenomenon compensable in tort, in this case nuisance, should encourage a resolution that keeps the matter out of litigation. Returning to the economically preferable approach of contract-based solutions, in a pollution scenario, the most efficient course of conduct will be where the polluter and the complainant reach an *ex ante* agreement regarding the level of harm the complainant is willing to sustain in return for the payment of money. This result, reached through cooperation and which avoids litigation, offers the lowest possible transaction costs, and the most efficient resolution.

¹³⁹ A tort rule that did not limit the complainant to one-time damages, and preclude future recovery by subsequent owners, *Boomer* suggests, would be inefficient. Subjecting the polluter to serial recoveries and indeterminate liability would constitute over-deterrence on the model of *Boomer*. Indeterminate liability would potentially fail to cap Atlantic Cement’s potential financial responsibility at a level that would permit it to continue to conduct business, and would thereby be inconsistent with the rule expressed in *Restatement (Second) of Torts* § 826(b), which permits the finding of nuisance even when the utility of the actor’s conduct outweighs the damage suffered by the complainant, so long as the damages are not set at a level that would prevent the defendant from continuation of its business. Additionally, an absence of the ‘one-time damages’ provision of *Boomer* would unjustly enrich the property owners, as they would recover first from *Atlantic Cement*, and then recover again by the sale of overvalued property, which is to say, property priced at a level that failed to take into account the chronic low-level future pollution.

¹⁴⁰ R H Coase, ‘The Problem of Social Cost’ (1960) 3 *Journal of Law and Economics* 1, reprinted in Saul Levmore (ed), *Foundations of Tort Law* (Foundation Press, 1994) 3 ff.

¹⁴¹ *Ibid.*

VI CONCLUSION

An optimal tort rule is just, moral and efficient. It advises those within its compass what is expected, what is discouraged, and the consequences of departure from the desirable. It does not compensate excessively but rather in proportion to the harm, and it does not undercompensate, as only through justifiable compensation is the rule's deterrent value most effective. It stands in the stead of *ex ante* agreements as to condoned or expected behavior in situations where contract would be impossible.¹⁴² At the same time, and by the same means, tort rules encourage safer behavior.

A so-called 'law of progress', affecting all disciplines from biology to law to history, is discussed by the influential R G Collingwood, who proposes that progress in history means simply that man builds his knowledge on the incidents of his experience and that of others.¹⁴³ Be it 'law' or not, a positivist social expectation of historical progress¹⁴⁴ has been foundational to the development of man's ethical and economic life. Success, social and political development are achieved when man has been informed as fully as possible of the past in such measure as will permit him to avoid its errors. Such change does not necessarily involve replacement of the bad with the good. As often, it will be the replacement of the good with the better.¹⁴⁵ A contemporary view that history reveals a linear progression propelled by good governance is urged by philosopher and neuroscientist Nayef Al-Rodhan, who argues that history may be defined as 'a durable progressive trajectory' in which the quality of life is

¹⁴² As Jules Coleman put it: '[t]ort law is not simply a necessary response to the impossibility of contract, but a genuine alternative to contract as a device for allocating risk': Jules L Coleman, *Risks and Wrongs* (Oxford University Press, 1992) 203.

¹⁴³ R G Collingwood, *The Idea of History* (Jan Van Der Dussen (ed)) (Oxford University Press, rev ed 1992) 322–3. This ageless improvement of tort law over time would be predicted by Socrates, in his constant references to the imperative of man's path of enhancing the life of the individual and the polity of man. 'Man does not and will not know that he participates in this progress': at 331.

¹⁴⁴ Many social historians urge that the path of history is one of steady improvement — herein of the modern idealists Hegel, Comte and others, first in Germany, then Europe, England and America, the disciples of 'progress' over 'providence' and of man's innate intellectual evolution. See Karl Lowith, *Meaning in History* (University of Chicago Press, 1949) 52, 67. French philosopher Auguste Comte, sometimes considered to be the founder of modern sociology, considered progress through gradual change inevitable, although he would have entrusted guidance of the polity to an elite of political experts, similar in ways to Plato's Philosopher-King. Marquis de Condorcet wrote on behalf of Philosophes that the 'perfectibility of man' is 'absolutely indefinite'. Thomas Jefferson subscribed to the Whiggish view advanced by leading British Unitarians in 'the indefinite perfectibility of man'. Only somewhat more modestly, DeToqueville wrote that in America at least, the concept of equality was strongly suggestive of a societal belief in the indefinite perfectibility of man.

¹⁴⁵ See Fukuyama, above n 4, 57, 58, 61, 145, referencing positivist arguments of, *inter alia*, Bernard Le Bovier de Fontenelle, Condorcet, Kant, Hobbes, Hegel and Locke.

premised on the human dignity needs of ‘reason, security, human rights, justice, and equality’.¹⁴⁶

Tort law has always been apiece with human optimism, and confidence in the capacity of man to improve himself and by so doing, improve society.¹⁴⁷ Tort law’s history, and ours, is that of the dialectic tension between self-aggrandizement¹⁴⁸ and self-abnegation. No other legal or ethical schema has so consistently hewn to the magisterial human experiment of justice, moderation, fairness, efficiency and equality in social groupings.

In this article are found multiple examples of both prophylaxes against and responses to wrongful infliction of harm to individuals or to the social collective. Such examples have ranged from the Babylonian response to the flooding of another’s land; to Socrates’ and Aristotle’s injunctions against unconsented-to taking; to Roman development of the law of nuisance; to Talmudic rules regarding waste. The development of tort rules over the ages reveals a theme of progress, rather than providence. It is seen that to the inexorable and permeating tort goals of fortifying morality and pursuing corrective justice can be fairly added objectives of economic efficiency, avoidance of waste, and cultivation of circumstances in which persons may preserve and protect their physical autonomy and their property — tort precepts by now evident in almost all cultures. As importantly, both corrective justice, with its moral underpinnings, and economic efficiency, have been routinely reconciled in greater or lesser harmony.¹⁴⁹ These dual objectives of tort law have been and will continue to be its elevation in comprehensiveness, its shedding of error, and its ongoing self-instruction guiding it to what is denominated sometimes as a ‘right fit’ in its time and culture.

¹⁴⁶ Nayef R F Al-Rodhan, *Sustainable History and the Dignity of Man: A Philosophy of History and Civilisational Triumph* (Lit Verlag, 2010). In ways comparable are Ghandi’s ‘Law of Progression’ and Alexis de Toqueville’s association of the American articulation of equality to a correlative belief in ‘the indefinite perfectibility of man’. Alexis de Tocqueville, ‘How Equality Suggests to the Americans the Indefinite Perfectibility of Man’ in *Democracy in America* (Vintage/Random House 1945) vol 2, ch 7.

¹⁴⁷ That modern man may overindulge this and other optimisms, see Robert Kurzban, *Why Everyone (Else) Is A Hypocrite* (Princeton University Press, 2010) 116–19, which may temper but does not vitiate this point.

¹⁴⁸ For thoughts representative of a laissez-faire social construct: ‘[i]t is an undeniable maxim that everyone by the light of nature and reason will do that which makes for him his greatest advantage’. Russell, above n 36, 624.

¹⁴⁹ Even avid exponents of an economic underpinning of nearly all tort rules recognize that such theories have prompted a ‘pushback’ from other social historians. Richard Posner wrote in the University of Chicago alumni magazine in 2011 that economic analysis of the law may have fallen victim of its own success, ‘becoming too esoteric, too narrow, too hermetic, too out of touch with the practices and institutions that it studies’: Peter Coy, *Will Success Spoil the Chicago School?* (7 June 2012) Bloomberg Businessweek <<http://www.businessweek.com/articles/2012-06-07/will-success-spoil-the-chicago-school>>. See also Coy, above n 114.

EXTRAJUDICIAL SPEECH AND THE PREJUDGMENT RULE: A REPLY TO BARTIE AND GAVA

ABSTRACT

The precise limits of the rule against prejudgment remain to be determined. It has recently been argued that the rule should be extended to prohibit extrajudicial statements on matters of law, as well as those of fact or evidence at issue in a particular matter. It is argued that this suggestion should be resisted, as neither the existing case law nor underlying principle support such an extension. Moreover, there are strong policy reasons for not doing so.

I INTRODUCTION

Susan Bartie and John Gava¹ challenge conventional understandings of the apprehended bias rule, and of the prejudgment doctrine in particular. They do so by arguing that this doctrine should be applied more broadly and that, for a variety of policy reasons, its scope should extend to a wide range of extrajudicial speech. Significantly, their argument forces a clearer conceptualisation of the bias rule and the legal values it protects.

Though far from new, extrajudicial commentary is a growing phenomenon. Judges of the superior courts are in increasing demand as conference speakers, often presenting keynote addresses, and a growing number of them have taken to the academic journals as well. Court websites often list lengthy catalogues of papers delivered by serving judges on a wide range of legal subjects to professional and academic audiences. Those papers range in nature from expressions of broad commentary about the legal system and the role of the courts to tightly argued expositions of a clear view as to the correct resolution of some question of specific legal doctrine.

Moreover, judges are increasingly willing to comment on, and express opinions about, 'hot' legal issues which are likely to be further argued in the nation's courts, perhaps even before their judicial selves. Given this context, Bartie and Gava warn against the possibility of a perception of prejudgment. They argue that 'ordinary human experience'² indicates that a position so clearly and publicly expressed will

* Associate Professor, Curtin Law School, Curtin University. My thanks to John Gava for encouraging me to write this response.

¹ Susan Bartie and John Gava, 'Some Problems with Extrajudicial Writing' (2012) 34 *Sydney Law Review* 637.

² *Ibid* 646.

be difficult for a judicial officer to put to one side when the very same legal issue subsequently falls to be determined in their courtroom.

Bartie and Gava suggest that much, if not all, extrajudicial speech is therefore ‘suspect’ in terms of displaying prejudice. They counsel a somewhat formalist remedy, a stony faced Sphinx like judicial silence in all fora other than duly delivered judgments. Judges, they seem to be saying, are best seen and not heard outside their courtrooms.

This article takes a different view. A close examination of the case law demonstrates that the prejudice rule has never extended to extrajudicial statements made about matters of law, as distinct from suggestions that particular questions of fact at issue in litigation in prospect have been predetermined. Moreover, while Bartie and Gava’s suggestion that strong expressions of extrajudicial views on matters of law or matters of fact are equally concerning has some initial appeal, this article argues that there are strong policy reasons for resisting their suggested extension of the bias rule.

II JUDGMENT AND PREJUDGMENT

Before engaging directly with the views expressed by Bartie and Gava, it is useful to consider exactly what is involved in judicial decision making. Judgment is a complex process. That said, at its simplest and most routine, it may be seen as being comprised of two key elements. On the one hand, courts must make relevant findings of fact on the basis of both admissible evidence and on a decision as to whether that evidence satisfies the applicable evidentiary burdens. The rules of evidence control this fact finding aspect of the judicial process.

On the other hand, the courts must also determine the applicable legal rules and principles. This involves interpreting relevant case law and statutory provisions to determine the intended judicial or legislative meaning. In relation to case law, questions of authority and persuasiveness must also be considered. Both statutory interpretation and the application of previous case law are complex processes, but neither is governed by the rules of evidence. Rather, judicial officers engage in a mix of inductive, deductive and analogical reasoning, with that mix varying from case to case. The ultimate decision of the court as to the applicable legal rules and principles is determined not by evidence, but by that reasoning process. It involves weighing up factors such as the precise statutory wording, the weight of case authority, and questions of underlying principle and policy.

At its most simplistic, the final determination of rights, obligations and liabilities in a judgment may be seen as a relatively straightforward process of deductive reasoning from the general premises provided by the applicable legal rules. Those rules, once having been determined on the basis of legal argument, are applied to the facts as found on the basis of the admissible evidence. The found facts form the minor premise of the deductive argument. The conclusion of the argument is the final judicial determination of liability or otherwise.

Necessarily, this is an idealised and overly simplistic description of judicial decision making. In many instances, a degree of judicial discretion will have to be exercised as the legal rules and the facts will admit of more than one conclusion. The idealised form of deductive reasoning will also be relatively rare in practice, since a judicial determination will only be called for where there is some degree of dispute between the parties as to either the interpretation of the applicable legal rules and principles or as to the primary facts to which those rules and principles are to be applied. Litigation will only ensue when there is dispute as to one or more of these matters.

Putting this admitted complexity to one side, there remain two essential elements in any judicial determination: the articulation of the relevant law and the finding of the relevant facts. Both may be contested, and both may be complex. Both may be matters to which a judicial officer might not bring a completely open and impartial mind. But they are very different functions, and ones typically performed at differing levels of the court hierarchy. Fact finding is principally a function performed at trial level, in the exercise of original jurisdiction, by a fact finder who has the benefit of hearing the witnesses and assessing their evidence firsthand. It is only in quite limited circumstances that questions of fact are directly redetermined at appellate levels. By contrast, questions of law, though routinely the subject of an initial determination at the trial level, are, equally routinely, the main subject of the exercise of appellate jurisdiction.

The question posed by Bartie and Gava is whether the prejudgment rule is equally applicable to these very different components of the judicial decision-making process. That question may itself be subdivided into two. First, what is the current law on the question? Second, is there a need for a reconsideration of that law?

III THE CASES ON PREJUDGMENT

To answer the first question, an examination of the decided cases is necessary. This examination yields two results. First, and consistently with the approach advocated by Bartie and Gava, judicial statements in the decided cases on prejudgment, or on the bias rule more generally, frequently make no overt distinction between these two separate aspects of the decision making process. Rather, those statements are couched in generality. In *Livesey v New South Wales Bar Association*³ the guiding principle was stated as follows:

a judge should not sit to hear a case if in all the circumstances the parties or the public might entertain a reasonable apprehension that he might not bring an impartial and unprejudiced mind to the resolution of the question involved in it.⁴

This is the classic statement of the rule against ‘apprehended’ or ‘ostensible’ bias, which may arise in a variety of ways, including prejudgment.⁵ On its face, the

³ (1983) 151 CLR 288 (*Livesey*).

⁴ *Ibid* 293–4.

⁵ *Webb v The Queen* (1994) 181 CLR 41, 74.

statement is broad enough to cover prejudgment in relation to issues of law and legal doctrine generally as well as prejudgment of the facts. More recently, in *Michael Wilson & Partners Limited v Nicholls*⁶ the High Court reaffirmed that:

the test to be applied in Australia in determining whether a judge is disqualified by reason of the appearance of bias (in this case, in the form of prejudgment) is whether a fair-minded lay observer might reasonably apprehend that the judge might not bring an impartial and unprejudiced mind to the resolution of the question the judge is required to decide.⁷

Again, the expression ‘the question the judge is required to decide’ is broad enough to capture a lack of impartiality with respect to both questions of fact and law.

However, a different picture emerges when one looks more closely at the facts of the cases within which these broad statements were made and the particular circumstances which were alleged to give rise to apprehensions of prejudgment. Consider a later statement of the High Court in the *Livesey* case, where Mason, Murphy, Brennan, Deane and Dawson JJ sought to apply the broad test laid down in that case to the circumstances of the matter before them:

a fair-minded observer might entertain a reasonable apprehension of bias by reason of prejudgment if a judge sits to hear a case at first instance after he has, in a previous case, expressed clear views either about *a question of fact which constitutes a live and significant issue in the subsequent case or about the credit of a witness whose evidence is of significance on such a question of fact*.⁸

First, note that the broad factual circumstance referred to here is a situation in which a judge, sitting at first instance, has previously heard another case. Then note that there are two things the judge may have done in that previous case, either of which might give rise to an apprehension of bias. Either the judge may have expressed, in that previous case, a clear view about a question *of fact* which is a live issue in the subsequent trial matter or, alternatively, the judge may have expressed a clear view as to the credibility of a witness whose evidence goes to this subsequently disputed question of fact.

These statements closely tracked the facts of the *Livesey* matter. In that seminal case the NSW Court of Appeal sat to determine an application that Livesey, a barrister, be struck off the roll of counsel on grounds of professional misconduct. Two of the three members of the Court of Appeal had previously determined proceedings arising out of the same factual matrix. Specifically, they had each held in those earlier proceedings that a key witness was untruthful and that Livesey had been a party to what was described as a ‘corrupt agreement’ or ‘conspiratorial arrangement’ with that key

⁶ (2011) 244 CLR 427.

⁷ *Ibid* 437 [31].

⁸ *Livesey* (1983) 151 CLR 288, 300 (emphasis added).

witness.⁹ This key witness was subsequently called by Livesey to retell her previously rejected version of events in the second set of proceedings. Thus, at the commencement of those proceedings involving Livesey, two of the three members of the Court of Appeal presiding had already made findings that the evidence of a key witness was false, and that Livesey knew this and was fully aware of, and indeed a party to, the ‘corrupt agreement’ or ‘conspiratorial arrangement’.¹⁰ As the presence or absence of this arrangement was foundational to the case against Livesey, it was not surprising that the High Court unanimously held that an apprehension of bias might arise in the circumstances and that the two members of the Court of Appeal involved in the previous proceedings should not have sat in the subsequent matter.

Other leading prejudgment cases follow a similar pattern. General statements of the rule against apprehended bias, and against prejudgment in particular, are couched in broad terms apparently capable of referring to statements of legal doctrine as well as matters of fact and credibility. But the second result of any detailed examination of the decided cases is that they focus firmly on the latter. In each instance, the relevant judicial remarks, subsequently found to provide a basis for a reasonable apprehension of bias, have been directed to issues of fact or credibility.

In another classic prejudgment case, *Vakauta v Kelly*,¹¹ the offending statements were made by a trial judge hearing a personal injuries matter without a jury. Liability having been conceded, the only issue to be determined was the quantum of damage, again a factual question. This was to be assessed on the basis of the evidence of medical witnesses for the parties. Before hearing their evidence, the trial judge made remarks strongly disparaging the credibility of three medical witnesses, famously describing them as an ‘unholy trinity’ who ‘think you can do a full week’s work without any arms or legs’ and whose ‘views are almost inevitably slanted in favour of the GIO [an insurance provider] by whom they have been retained, consciously or unconsciously’.¹² These remarks were held to constitute prejudgment of the credibility of those medical witnesses, and hence of the issue of fact their evidence addressed.¹³

The High Court recognised the reality that a trial judge may well be confronted by witnesses who are repeat players and whose evidence, in an adversarial context, may tend to favour the side that has retained them. Brennan, Deane & Gaudron JJ observed that the rule against prejudgment ‘must be observed in the real world of actual litigation’.¹⁴ Their Honours continued:

⁹ Ibid 297.

¹⁰ Ibid.

¹¹ (1989) 167 CLR 568.

¹² Ibid 572 (Brennan, Deane and Gaudron JJ).

¹³ Ibid 573–4 (Brennan, Deane and Gaudron JJ). The conclusion that an apprehension of bias might reasonably arise was reinforced by additional remarks made by the trial judge in the course of the reserved judgment.

¹⁴ Ibid 570.

That requirement will not be infringed merely because a judge carries with him or her the knowledge that some medical witnesses, who are regularly called to give evidence on behalf of particular classes of plaintiffs (eg members of a particular trade union), are likely to be less sceptical of a plaintiff's claims and less optimistic in their prognosis of the extent of future recovery than are other medical witnesses who are regularly called to give evidence on behalf of particular classes of defendants (eg those whose liability is covered by a particular insurer).¹⁵

So, an open mind does not mean a completely blank one — a point that will be worth returning to in the context of extrajudicial speech. Importantly for present purposes, however, the impugned remarks of the trial judge focused entirely on the credibility of the witnesses, and hence on the determination of the questions of fact to which their subsequent evidence related.

In the other leading prejudgment case, *Minister for Immigration and Multicultural Affairs v Jia Legeng*¹⁶ the High Court decided that no apprehension of bias arose in the context of remarks made by the then Minister for Immigration to a talkback radio host. Again, those remarks, albeit couched in somewhat general terms, were made in direct response to decisions made by the Administrative Appeals Tribunal in relation to a particular person; decisions which the Minister would shortly be called upon to respond to. Two points are relevant. First, the fact that the impugned remarks were made by a Minister 'who not only has general accountability to the electorate and to Parliament, but who ... is made subject to a specific form of parliamentary accountability'¹⁷ rather than a judicial officer, was central to the finding that no apprehension of bias would arise from those remarks:

The position of the Minister is substantially different from that of a judge, or quasi-judicial officer, adjudicating in adversarial litigation. It would be wrong to apply to his conduct the standards of detachment which apply to judicial officers or jurors.¹⁸

Second, the question allegedly the subject of prejudgment by the Minister, whether or not Mr Jia was of 'good character' for the purposes of s 501,¹⁹ was again a question of fact, or at most a mixed question of fact and law. It was the discussion of Mr Jia's particular case by the Minister which triggered the ultimately unsuccessful allegations of apprehended (and actual) bias.

There is a subclass of prejudgment cases which further illustrates the point. This subclass focuses on persistent rudeness, interjection or harassment by a trial judge or

¹⁵ Ibid 570–1.

¹⁶ (2001) 205 CLR 507.

¹⁷ Ibid 539.

¹⁸ Ibid 539 (Gleeson CJ and Gummow J).

¹⁹ *Migration Act 1958* (Cth) s 501. The section provides for refusal or cancellation of visas on character grounds.

tribunal member, again suggesting that the credibility of a witness, a party, or their representative has been adversely predetermined. Examples in this category include *Damjanovic v Sharpe Hume & Co*²⁰ and *Re Refugee Review Tribunal; Ex parte H*.²¹

In *Damjanovic*, an otherwise unrepresented plaintiff sought assistance in the District Court from a friend who was not legally qualified. The District Court judge only reluctantly acceded to this request, at the same time ruling that six separate matters involving the same plaintiff be heard together, with the evidence in each matter to be evidence in each of the others. The course of the trials over several days was marked by ‘ever increasing’²² criticism of the unrepresented applicant and his lay representative, threatening statements as to the likely outcomes of the litigation, abruptness, rudeness, interruptions and sarcastic remarks, apparently uneven treatment between the parties, and more. Events culminated in the judge determining one of the matters before the evidence was complete in all of them, despite having ruled earlier that they were to be tried together.

In *Ex parte H*, a claimant for refugee status was subjected to highly sceptical questioning from the Refugee Review Tribunal member. The transcript of the tribunal hearing showed constant interruptions of the evidence the applicant was attempting to give, and repeated statements by the Tribunal member that they did not believe the applicant. Gleeson CJ, Gaudron and Gummow JJ considered that:

a fair-minded lay observer or a properly informed lay person, in our view, might well infer, from the constant interruptions of the male prosecutor’s evidence and the constant challenges to his truthfulness and to the plausibility of his account of events, that there was nothing he could say or do to change the Tribunal’s preconceived view that he had fabricated his account of the events upon which he based his application for a protection visa.²³

It is unsurprising that in both matters, it was held that a reasonable apprehension of bias, and specifically of prejudgment, might well have arisen. The relevant point for present purposes is that in each of these cases, it is clear that the comments which gave rise to the apprehension of bias were directed to the credibility of the witness, and hence to the version of the facts they were seeking to put forward. *Damjanovic* was a case being determined at the trial level when the impugned remarks were made; in *Ex parte H*, the Refugee Review Tribunal was exercising a merits review jurisdiction, and hence re-deciding the relevant questions of fact for itself.

More recent decisions are similarly instructive. In *British American Tobacco Australia Services Limited v Laurie*,²⁴ the High Court held, by 3:2 majority, that

²⁰ [2001] NSWCA 407 (21 November 2001) (Mason P, Sheller JA and Rolfe AJA) (*‘Damjanovic’*).

²¹ (2001) 179 ALR 425 (*‘Ex parte H’*).

²² *Damjanovic* [2001] NSWCA 407 (21 November 2001) [158].

²³ *Ex parte H* (2001) 179 ALR 425, 435 [32].

²⁴ (2011) 242 CLR 283.

a judge (once again a trial judge) was disqualified from hearing a matter when he had heard and determined very similar questions of fact and credibility involving the same respondent at an interlocutory stage of a different matter some three years earlier. The dissenting judges, French CJ and Gummow J, based their dissent on the guarded and provisional language used by the judge in that earlier interlocutory decision in relation to those questions of fact and credibility. The majority, applying *Livesey*, considered the language used by the judge to be sufficiently robust as to give rise to the possibility of an apprehension that the judge might not be able to put to one side those earlier opinions as to fact and credibility.²⁵ Given the narrow majority, and the fact that the impugned remarks took place during the hearing of another matter, the case can be considered to be at the outer limits of the factual circumstances that will found an apprehension of bias via prejudgment.

In *Michael Wilson & Partners Limited v Nicholls*,²⁶ the hearing of numerous ex parte applications was held not to give rise to any apprehension of bias on the part of the judge in question, therefore not disqualifying them from hearing the subsequent substantive matters. In the view of the High Court, the factual questions to be determined in the interlocutory applications were quite separate, and did not involve the prejudgment of any relevant questions of fact or credibility arising in the later litigation.²⁷ The focus upon factual questions was clear.

What conclusions can we draw from this? In essence, the decided prejudgment cases focus very much on predetermination of factual issues, whether directly or via prejudgment of the credibility of witnesses to those facts. None of these cases would appear to involve a scenario where the impugned judicial remarks focused upon questions of legal doctrine. Broad statements of the prejudgment rule may appear consistent with the broader scope for that rule advocated by Bartie and Gava, but the decided cases fall within a much narrower compass. This in turn refines our understanding of the bias rule. Particularly in relation to prejudgment, that rule has never been so broad as to refer to any matter that might incline a judge to decide a matter one way or the other; rather, it has focused more narrowly and specifically on prejudgment of matters of fact or credibility.

There are occasional decisions from other jurisdictions that do not fit this pattern. Bartie and Gava identify a decision of the Scottish High Court of Justiciary, *Hoekstra v Her Majesty's Advocate [No 3]*,²⁸ that falls into this exceptional category. In that case it was held that extrajudicial criticism of the European Convention on Human Rights and its adoption in Scots law could suggest bias that would render any trial adjudicated by that judicial author unfair under Article 6 of that Convention. This result is indeed consistent with the broader position advocated by Bartie and Gava. However, the decision does not seem to have attracted a broad following. The UK Court of Appeal declined an opportunity to follow a similar path, on very similar

²⁵ Ibid 333 [145].

²⁶ (2011) 244 CLR 427.

²⁷ Ibid 448 [73] (Gummow ACJ, Hayne, Crennan and Bell JJ).

²⁸ 2000 SLT 605.

facts, a year later in *R v Spear*.²⁹ In that case, the defendants to a court martial had submitted that their trials violated Article 6 of the European Convention. The Court of Appeal rejected this argument, and also dismissed a suggestion that remarks made by Laws LJ in the course of its hearing, which might have been seen as critical of the Convention, gave rise to any appearance of bias on his part.

In *Locabail (UK) Ltd v Bayfield Properties Ltd*³⁰ the Court of Appeal summarised the position as follows:

It would be dangerous and futile to attempt to define or list the factors which may or may not give rise to a real danger of bias. Everything will depend on the facts, which may include the nature of the issue to be decided. We cannot, however, conceive of circumstances in which an objection could be soundly based on the religion, ethnic or national origin, gender, age, class, means or sexual orientation of the judge. Nor, at any rate ordinarily, could an objection be soundly based on the judge's social or educational or service or employment background or history, nor that of any member of the judge's family; or previous political associations; or membership of social or sporting or charitable bodies; or Masonic associations; or previous judicial decisions; or extra-curricular utterances (whether in text books, lectures, speeches, articles, interviews, reports or responses to consultation papers).³¹

The conclusion to be drawn from this examination of the decided cases is that the current law does not support the position advocated by Bartie and Gava. It would seem a significant extension of the existing rule against prejudgment to sanction judicial commentary that limits itself to debate about relevant legal principle. This article now turns to the second question at issue, the one at the heart of Bartie and Gava's position; is there a persuasive argument in policy and principle terms that the prejudgment rule be widened from its current scope to include a much broader range of extrajudicial statements?

IV EXTRAJUDICIAL COMMENTARY

Extrajudicial speech comes in a number of forms. Some of these are quite discursive; others resemble some elements of a written judgment in their discussion of the decided cases. Quite appropriately, such utterances rarely if ever focus on disputed matters of fact in pending litigation. However, as Bartie and Gava point out, they may express clear views on a point of law that may well turn out to be decisive in subsequent litigation. It is this observation which makes their suggestion that such extrajudicial statements ought to be regarded as falling within an extended conception of objectionable 'prejudgment' initially attractive.

²⁹ [2001] QB 804 (Laws LJ, Holman and Goldring JJ).

³⁰ [2000] 1 QB 451.

³¹ *Ibid* 480 [25] (emphasis added).

An example of extrajudicial speech not dissimilar to a judgment is the article by the Hon David Ipp, dealing with the proper legal test for the tort of malicious prosecution.³² This article is, as Bartie and Gava observe, ‘the product of careful consideration of all the relevant authorities and is highly polished and tightly reasoned’.³³ Essentially, the piece considers a conflict of authorities and comes down, quite clearly, on one side of the debate. Although policy considerations are mentioned, the burden of the argument falls simply upon the precedential weight of the conflicting authorities. It reads very much like a section of a written judgment wherein the judicial officer carefully reviews the authorities bearing upon the point at issue, displaying a sensitivity to the factual issues at stake in the precedent cases and balancing the weight of the judicial rulings and dicta contained therein. As Bartie and Gava comment, the piece ‘could easily be transposed into a judgment with very little alteration’.³⁴

One might observe, however, that the hypothetical transposition might just as easily have occurred in the opposite direction. This observation weighs against the broadening of the bias rule advocated by Bartie and Gava, as it reveals the policy implications of such a broadening to be unacceptable. If academic extrajudicial speech, such as the Ipp article, were to be considered as falling within the ambit of the apprehended bias rule, then it would be difficult to exclude any previous judgments delivered by a judicial officer which dealt with the same area of law from also falling within that rule. Both could be considered to display ‘prejudgment’ of legal issues in the extended sense advocated by Bartie and Gava. As they correctly observe, there is very little that distinguishes the kind of academic discussion of relevant authorities, leading to a reasoned conclusion on a point of law, displayed in this article by Ipp from the kind of judicial reasoning displayed in previous cases. If the one betrays a prejudgment of questions to be determined in subsequent litigation, it is difficult to see that the other does not. The apprehended bias rule should apply, or not apply, equally to both varieties of judicial utterance.

But to cast the net so broadly as to include both would surely make the court system entirely unworkable. A great deal of the work of judges, magistrates and tribunal members involves routine determinations in high volume jurisdictions, such as personal injuries, workers compensation, refugee status determinations and so on, with serial players repeatedly litigating the same key legal issues. Disqualification on the basis that a judge had previously heard a matter raising similar legal issues, and hence was on record expressing a view as to the correct resolution of those issues, would be quite unworkable.

Bartie and Gava suggest that the analogy between extrajudicial speech and previous judgments is not a strong one. They observe that:

³² D A Ipp, ‘Must the Prosecutor Believe that the Accused is Guilty? Or, Was Sir Frederick Jordan Being Recalcitrant?’ (2005) 79 *Australian Law Journal* 233.

³³ Bartie and Gava, above n 1, 648.

³⁴ *Ibid* 649.

The history and tradition of the common law requires judges to commit to certain doctrinal positions in judgments. These commitments are legitimate and are binding in future matters because they are grounded by a fair hearing of the arguments and issues raised by counsel.³⁵

It is difficult to see how this makes the case that previous judicial decisions are any less likely to embody 'prejudgment' than extrajudicial speech. If anything, the opposite conclusion might easily be drawn. First, conclusions reached in actual judgments are necessarily the weight of considered reflection on the matter, with the benefit of argument from opposing counsel, and in the light of the particular facts of the matter at hand. That reflection is all the more likely to involve commitment from the judicial officer to the position finally adopted, given the significance of that position for the parties to the litigation. A judicial decision is anything but 'academic', in the sense that it has direct and immediate consequences for the parties to the litigation. It is necessarily the result of the most careful and considered decision by the judicial officer in question. Moreover, considerations of precedent, consistency and comity, none of which apply in the same way to extrajudicial speech, each place pressure on judges to decide new cases consistently with their own previous decisions. Indeed, the argument put by Ipp in relation to the tort of malicious prosecution is at its heart a complaint that insufficient weight had been given to this need to adhere to relevant and applicable precedent.

There are other, more discursive, forms of extrajudicial speech which are not so closely analogous to written judgments. Bartie and Gava's second example is a well-known extrajudicial piece written by the Hon Ian David Francis Callinan, and published some 13 months prior to his retirement as a judge of the High Court.³⁶ The article, published in *Quadrant*, is the text of an address entitled *International Law and Australian Sovereignty*. In it, Callinan expresses, in quite broad terms, a series of views as to the influence of international law. Overall, those views are hostile to international law and what Callinan sees as its growing influence. This influence is argued to be both anti-democratic and generally malign. International instruments are criticised as being difficult to adapt to new circumstances and are compared unfavourably with common law protections of human rights, particularly in the criminal justice process.

These views are all very general in their nature and fall well short of expressing anything like a decided view on any issue which Callinan might conceivably have been called to decide. A view that the external affairs power may have been over-extended does not translate to a prejudgment as to the validity of a particular piece of legislation made in exercise of that power, any more than a piece of judicial obiter on that topic would do so. This is all the more so, given the weight of pre-existing authority in the Court. At the most, it expresses a degree of scepticism as to the employment of international instruments, and a robust reminder of the conventional wisdom that they do not, absent incorporation, become part of Australian law. These are views that a judge is entitled to hold, without being considered to have prejudged litigation in which they may be tangentially raised.

³⁵ Ibid 655.

³⁶ I D F Callinan, 'International Law and Australian Sovereignty' (2005) 49 *Quadrant* 9.

In a similar vein is the discussion by Bartie and Gava of the UK decision in *Pepper v Hart*³⁷ and the subsequent criticism of that decision made extrajudicially by Lord Steyn.³⁸

In that celebrated decision, the UK House of Lords decided that it was permissible to have recourse to extrinsic materials in order to assist in the clarification of legislative intent, a practice which has become relatively non-controversial in Australia. In the UK, however, the decision was the subject of strong criticism. It was said to be contrary to existing precedent and, in the view of its critics, to constitutional principle, as well as potentially adding to the length, complexity and cost of litigation. Lord Steyn was a strong voice amongst those judicial and academic critics, though far from alone, and wrote a number of pieces developing his concerns, which led ultimately to his concluding ‘... it follows that *Pepper v Hart* is not good law’.³⁹

It is this conclusion, based upon what Bartie and Gava describe as ‘a number of pages of carefully reasoned argument’ during which Lord Steyn ‘states his position as a clearly delineated legal rule’⁴⁰ that they see as exemplifying prejudgment.

But let us consider again the issue in *Pepper*. The issue to which Lord Steyn directed his criticism was the use of extrinsic materials. No court will be called upon to rule on that issue in abstraction. *Pepper* itself illustrates the point, where the actual question before the House of Lords was whether some schoolteachers were to continue to receive the benefit of perks, in the form of cheap schooling for their own children at a greatly reduced tax rate, where the legislation apparently removed this long standing ‘fringe benefit’.

In order to determine that question, the Lords had necessarily to determine the intended meaning and purpose of the legislation in question. It was argued that statements of the relevant Minister cast light on that issue. Controversially, the Lords decided that it was indeed permissible to refer to those statements, and did so. Accepting for the purpose of argument that the admitted material was indeed determinative of the statutory interpretation issue, and thus ultimately dispositive of the litigation, the two issues remain entirely separate in nature. They were ultimately connected in this litigation, but only by a series of argumentative steps. In reaching its ultimate decision in *Pepper* the House of Lords had at least to do the following:

- (i) reach factual conclusions, on the basis of the admissible evidence (in practice, the factual disputes are usually settled in the lower courts);
- (ii) identify interpretive issues arising in the applicable legislative provisions;

³⁷ [1993] AC 593 (*Pepper*).

³⁸ Johan Steyn, ‘*Pepper v Hart*; A Re-examination’ (2001) 21 *Oxford Journal of Legal Studies* 59.

³⁹ *Ibid* 70.

⁴⁰ Bartie and Gava, above n 1, 647.

- (iii) decide whether it was appropriate to admit extrinsic material and have regard to it in order to resolve those interpretive issues;
- (iv) determine the meaning of the applicable legislative provisions, in the light of the admitted extrinsic material; and
- (v) determine the ultimate question of tax liability for the schoolteachers by applying the relevant statutory provisions, as interpreted with the assistance of the admitted extrinsic material, to the facts as found.

The issue which so animated Lord Steyn in his subsequent commentaries was the third of these — whether extrinsic material should be admitted into evidence to assist in the determination of the legislative meaning. It was, and is, a general issue of procedure that may, or may not, arise in the course of any litigation involving the application of legislation to a particular set of facts. Arguably, it will seldom be determinative of the litigation in question. Even in those instances where it is determinative, it remains an analytically distinct issue from the actual question of rights, obligations and liabilities before the court. It is one thing to say that a judge, in this instance Lord Steyn, has expressed strong views on this issue; it is quite another to say that he has prejudged the question of actual liability. It is utterances directed to the latter question which most clearly give rise to apprehensions of prejudgment. If, for example, those members of the House of Lords determining *Pepper* were to announce, prior to hearing the case, firm views as to the liability or otherwise of the schoolteachers to pay the additional tax, then that would indeed smack strongly of prejudgment.

V CONCLUSIONS

It has been argued above that even quite strongly expressed views as to legal doctrine that may ultimately prove dispositive of a particular matter should not be seen as raising legally objectionable apprehensions of bias in relation to that, or any other, particular matter. The key reason is that such statements are simply not directed to any particular matter. Rather, they are expressions of broad general views. It is argued here that it is expressions of opinion as to the *facts* of a matter, prior to the reception of all the evidence on that matter that rightly raise concerns as to prejudgment.

The law of procedural fairness has always been particularised in the sense that it is directed to the provision of a fair and impartial hearing to an *individual* in the context of a *particular*, usually adverse, decision to be made with respect to that individual. Thus, the right to a ‘fair’ hearing, the right to be heard, has never developed into a broad right of citizens to be consulted on matters of governmental policy. Similarly, the right to an ‘impartial’ hearing is a particularised one. The impartiality which the rule against bias protects equates to a right that a particular person’s guilt or liability be determined only on the basis of admissible evidence,⁴¹ and not be prejudged prior

⁴¹ Accepting, of course, that the rules of evidence are not directly applicable to non-judicial decision makers.

to that evidence being heard. It is no breach of that principle that a judge has a strong view as to the *law* to be applied to the case prior to hearing it, so long as an open mind as to the facts is retained until all the evidence is in.

As noted above, a contrary view would not only rule out extrajudicial speech, but would also necessarily see previous judicial decisions as indicative of prejudgment. The closer the facts, the stronger that indication will be. This, after all, is the doctrine of precedent. Those previously decided cases give a much stronger indication of the way subsequent ‘similar’ matters are likely to be decided than can be gleaned from comparatively abstract and generalised academic writing.

It is useful to also consider the other key rationale for procedural fairness doctrines. Neither the hearing rule nor the bias rule is simply about providing ‘fairness’ to an individual, important though that is. Rather, these doctrines are also instrumental, in that they aim to achieve greater accuracy in decision making by ensuring that a decision maker is more fully informed when they come to make their decision. This rationale is particularly pertinent when it comes to prejudgment of the facts of a matter, since these facts are usually quite outside the private knowledge of the judicial officer in question. Hence, the insistence of the prejudgment rule that a judge keep an open mind and not make any determination of the facts of a particular matter until all the evidence that might assist in that determination has been duly received in court. It is also in this context that the rules of evidence become significant, designed as they are to ensure that, as much as possible, the determination of the facts is made only upon the basis of the most reliable evidence.

By contrast, judicial officers are fully expected to have expert legal knowledge, and are appointed to their courts on that basis. The rules of evidence do not apply to the acquisition of that knowledge. Whilst a court will certainly hope for the assistance of counsel in determining contested legal issues, it is unlikely to be starting from a blank slate.

In conclusion, let us consider Bartie and Gava’s suggested remedy. They counsel ‘judicial silence’,⁴² but this can only mean one of two things. With respect, both are equally unsatisfactory.

The first possible meaning is that judges should not have opinions — that their minds should be empty. This of course is neither possible nor desirable. It is inevitable that judges will be aware of doctrinal issues and the competing arguments that surround them. They will have considered such matters from their formative days as students, then as barristers and very commonly in previous cases they have decided. They are appointed to the bench principally on the basis of their years of accumulated knowledge and experience of the law. To suggest that all this is somehow to be put aside is implausible. Neither would it be desirable. If a judge is finally called upon to decide a contentious point, it is surely desirable that they are able to bring as much well-informed reflection as possible to the arguments put to them by counsel.

⁴² Bartie and Gava, above n 1, 639.

It would not be a good thing to require judges to be entirely free of preconception or education, even if it were possible and it would hardly be to the benefit of litigants. For this reason, the bias rule has never required that a judge bring an empty mind to matters they are required to determine, but only that they bring an open mind, in the sense of a mind being open to persuasion. An impartial mind is not an empty one.

More probably, what Bartie and Gava are suggesting is that judges should keep the opinions that they will inevitably possess strictly to themselves. But this would be a superficial change only. Silence might remove the appearance of predisposition on legal questions, but could not remove its actuality. Such concealment would not serve the interests of justice or of litigants. Far better for judicial views to be on the record, so that counsel can be alerted to them and can address them specifically. It is submitted that fairness requires no less.

FRANCHISING: A HONEY POT IN A BEAR TRAP

ABSTRACT

A franchisee's business is like a honey pot inside a loaded bear trap. This article explores the laws that purport to protect the interests of franchisees (namely contract law, the *Competition and Consumer Act 2010* (Cth) and the Franchising Code of Conduct) and identifies why those laws are doomed to failure if the franchisor becomes insolvent. The example of failed franchisor Angus & Robertson is used to explain the franchisees' predicament in the face of the responsibilities imposed on insolvency practitioners by the *Corporations Act 2001* (Cth). The article concludes that piecemeal reform of consumer protection legislation can never result in fair franchising. An approach that includes amendments to insolvency policy and regulation is required.

I INTRODUCTION

Business format franchising has been described as 'a form of "honey trap" into which inexperienced franchisees [are] lured by promises of success'.¹ On investigating a franchise the franchisees and their advisors test the honey, but seldom notice, examine or understand the power of the bear trap that surrounds it. That bear trap is franchisor insolvency. The power to trip this bear trap sits with the franchisor and the franchisor's creditors. Once tripped, all franchisees are caught within its teeth. Most are unable to escape until the franchisor's administrator or liquidator frees them. Notwithstanding this flaw, franchising remains an important part of Australia's economy.² In 2011, the Parliament of South Australia heard that

approximately \$180 billion nationally is devoted to franchise expenditure ... the numbers are actually amazing. There are 1270 different franchise operations existing in Australia, and 670 of those operate in South Australia alone. Nationally, franchising employs 775 000 people.³

* Jenny Buchan, Senior Lecturer, School of Taxation and Business Law, Australian School of Business, University of New South Wales.

¹ Economic and Finance Committee, Parliament of South Australia, *Franchises* (2008) 18.

² Jenny Buchan, *Franchisees as Consumers: Benchmarks, Perspectives and Consequences* (Springer, 2013) 16–17.

³ South Australia, *Parliamentary Debates*, House of Assembly, 13 September 2011, 4878 (Steven Griffiths). Based on South Australia having 7.2 per cent of Australia's population franchising would provide work for 55 800 South Australians.

The law is playing a catch-up game with this evolving and nuanced business model. As South Australia considers enacting an industry code⁴ for the franchise sector⁵ it is timely to reach into the honey pot in an effort to determine the size and power of the bear trap it inhabits. How strong are its jaws? Do they need to be weakened? If so, how might this be achieved?

This article explores the fraught pathway of a franchisor's insolvency through the experiences of franchisees, primarily those connected with failed franchisor Angus & Robertson ('A & R'). It also suggests how franchisees can respond to the failure of their counterparty within the boundaries of the current law. Concluding that the current situation is unsatisfactory, it suggests where solutions lie. Part II introduces the concept of franchisor failure: how often do franchisors fail, and why? Part III then identifies indicators of a company's insolvency, using this as a background to investigate the failure of A & R in Part IV. Part V examines the precarious position of franchisees in circumstances of franchisor failure, in light of the current legal structure. Part VI outlines *ex ante* and *ex post* courses of action available to franchisees to anticipate and mitigate their losses within the current law. Part VII identifies potential solutions, and Part VIII concludes.

II FAILURE IN FRANCHISING

This article is concerned with the failure of franchisor insolvency, including administration and liquidation. The law does not provide a predictable path through franchisor failure for franchisees. The South Australian Franchise Review ('SA Review') accepted that

many franchisees entering franchises are not in a position to anticipate the difficulties they may face as a result of the failure of their franchisor. ... The provision of ... information [about consequences of franchisor failure on a generic level is] insufficient to address the current regulatory gap.⁶

The specific structure of the franchisor's network, its level of debt, the availability of suitable buyers for the failing franchisor, and each individual franchisee's personal and financial resilience and negotiating ability influence how individual franchisees emerge from their franchisor's failure. 'Some can draw from a deep well of prior experience ... and a strong personal support network; some [should] only attribute the survival of their business to luck or its demise to bad luck'.⁷ But, is franchisor failure a big enough issue to merit concern about its consequences?

⁴ Evidence to Estimates Committee A, Parliament of South Australia, Adelaide, 22 June 2012, 195 (Tom Koutsantonis).

⁵ Pursuant to powers granted under the *Small Business Commissioner Act 2011* (SA) s 14.

⁶ Economic and Finance Committee, above n 1, 37–8.

⁷ Jennifer Mary Buchan, *Franchisor Failure; An Assessment of the Adequacy of Regulatory Response* (PhD Thesis, Queensland University of Technology, 2010).

A How Often Do Franchises Fail?

Before considering the consequences of franchisor failure it is useful to understand the magnitude of the issue. As long as 35 years ago researchers in the United States ('US') observed that many franchises were failing, identifying '54 entire restaurant franchise systems that turned "belly up" over a two-year period'.⁸ Other American researchers noted that out of an estimated population of 2177 franchisors in 1986: '[a] total of 104 franchisors operating 5423 outlets failed [the following year] ... The annual volume of sales represented by those failed firms was [US]\$1.7 billion, of which the franchisee-owned portion was [US]\$1.5 billion'.⁹ More recently, Rozenn Perrigot and Gérard Cliquet studied 952 French franchising networks, and found that during the 10 year period from 1992–2002, nearly 58 per cent of franchisors failed.¹⁰

Australian franchisors fail too. 'The 1999 Australian Franchising Yearbook and Directory listed 347 franchisors. Of these, 251 (72 per cent) were no longer franchising'¹¹ by 2011. There were 1100 franchisors trading in Australia in 2008 and by 2010 a total of '56 franchise systems had ceased operating and a further 88 ceased franchising'.¹² This was 13 per cent in two years — more than one franchisor in five over two years. By 2012 a further 48 franchisors had stopped franchising in Australia.¹³ The Australian numbers include franchisors that ceased franchising but possibly remained in business, and others that failed, like Kleenmaid (15 franchisees),¹⁴ Kleins Jewellery

⁸ Shelby D Hunt, 'Franchising: Promises, Problems, Prospects' (1977) 53(3) *Journal of Retailing* 71, 75, citing Select Committee on Small Business, United States Senate, *The Economic Effects of Franchising* (1971) 93.

⁹ Roger D Blair and Francine Lafontaine, *The Economics of Franchising* (Cambridge University Press, 2005) 272; see also Andrew KostECKA, *Franchising in the Economy* (United States Department of Commerce, 1988) 12.

¹⁰ Rozenn Perrigot and Gérard Cliquet, 'Survival of Franchising Networks in France from 1992 to 2002' (Paper presented at the 18th Annual International Society of Franchising Conference, Las Vegas, Nevada, 6–7 March 2004).

¹¹ Jenny Buchan, Charles Qu and Lorelle Frazer, 'Protecting Franchisees from their Franchisor's Impending Failure: A Way Forward for Consumer Protection Regulators Using Indicators?' (Paper presented at The First International Conference on Comparative Law and Global Common Law: Exchange and Integration of the Contemporary Legal Systems, College of Comparative Law, China University of Political Science and Law, Beijing, 24–25 September 2011); Blair and Lafontaine, above n 9, 44, 272; Francine Lafontaine and K L Shaw, 'Franchising Growth and Franchisor Entry and Exit in the US Market: Myth and Reality' (1998) 13(2) *Journal of Business Venturing* 95, in Frank Hoy and John Stanworth (eds), *Franchising: An International Perspective* (Routledge, 2003) 163–4; Perrigot and Cliquet, above n 10.

¹² Buchan, Qu and Frazer, above n 11.

¹³ Lorelle Frazer, Scott Weaven and Kelli Bodey, *Franchising Australia 2012* (Griffith University, 2012).

¹⁴ The Kleenmaid franchisees paid initial fees of \$60 000 to \$120 000 each plus setup costs. They subleased premises from the franchisor's associate, Kleenmaid Property: Deloitte, 'The Kleenmaid Group (Administrators Appointed) Administrators' Report to Creditors Pursuant to s 439A *Corporations Act 2001* (Cth)' (Administrators' Report, 14 May 2009) 12.

(134 franchisees),¹⁵ Traveland (270 franchisees), Beach House Group (22 franchisees), Healthzone Limited (80 franchisees), Refund Home Loans (320 franchisees), Tyrecorp (33 franchisees), Darrell Lea (69 company-owned and franchised stores), Master Education Services (well known for its test case)¹⁶ and A & R (48 franchisees). A franchisee that is receiving negligible assistance from a franchisor is likely to quit the system before the franchisor fails. The fate of each franchisee that is still in the system when the franchisor fails will depend on many variables: some also fail, others rebrand as a franchisee of another system, and yet others become independent businesses.

Each franchisor has between one and thousands of franchisees. ‘The proportion of franchisors to franchisees [averages] 1:60 in Australia’.¹⁷ This ratio lends weight to the proposition that one franchisor failing has a far greater economic impact than one franchisee failing, and indeed, it suggests that the issue is serious enough to investigate its consequences. In economic terms franchising is a classic example of market failure. Franchisors pass risk and the consequences of franchisor failure on to franchisees. The externalities¹⁸ that manifest themselves when franchisors fail are not costed into the franchise model.

Urban myth is responsible for the claim that franchisees cause their franchisor’s failure. It is thus important to probe the myth. Maybe franchisees get what they deserve?

B *Why Do Franchisors Fail?*

In 1991, when business format franchising was not yet regulated in this country, Australia’s Franchising Task Force attributed franchisor failure to a combination of: under-capitalization of the franchisor, too-rapid expansion of the franchise system, poor product or service, poor franchisee selection, franchisor greed, external factors, devaluation of the Australian dollar, an increase in import duties, the withdrawal of an important source of products, an aggressive and cheaper competitor, and severe downturn in the economy.¹⁹ In the US, Cross saw ‘[f]ailure as a result of “franchising-related” factors as falling into five key categories: business fraud, intra-system competition, involving franchise outlets being located too close,

¹⁵ James Thomson, *Kleins Collapse Turns Nasty* (30 November 2012) Smart Company <<http://www.smartcompany.com.au/growth/economy/973-kleins-collapse-turns-nasty.html>>; see also Birkett Long Solicitors, *What Happens if a Franchisor Fails?* (28 September 2009) <http://www.birkettlong.co.uk/site/library/legalnews/what_happens_if_a_franchisor_fails.html>.

¹⁶ *Master Education Services Pty Limited v Ketchell* (2008) 236 CLR 101; *Ketchell v Master of Education Services Pty Ltd* (2007) 226 FLR 169.

¹⁷ Buchan, *Franchisees as Consumers*, above n 2, 16.

¹⁸ Externalities have been described as ‘items missing a price tag’: Ian Fletcher, ‘The Theory That’s Killing America’s Economy — And Why It’s Wrong’, *Huffington Post* (online), 4 July 2011 <http://www.huffingtonpost.com/ian-fletcher/the-theory-thats-killing-_b_846452.html>.

¹⁹ Franchising Task Force, *Franchising Task Force Final Report: Report to the Minister for Small Business and Customs The Hon David Beddall MP* (The Task Force, 1991).

insufficient support of franchisees, poor franchisee screening, [and] persistent franchisor-franchisee conflict'.²⁰ These are in addition to the traditional causes for small and medium-sized-enterprise ('SME') failure identified by Cross as being 'generic' causes which should be ameliorated by franchising: undercapitalization, absence of economies of scale, lack of business acumen and inability to survive intense competition in sectors where barriers to entry are low.²¹

The complexity of today's franchise environment suggests that additional threats to the security of franchisees' interests are emerging. Particularly problematic are: the failure of the franchisor's parent company, franchisor ownership by venture capitalists, the absence of governance duties being owed by franchisors to franchisees,²² the awkward accommodation of franchisees under Australia's insolvency law, and strategic insolvency. For Traveland, for example, failure of its parent company Ansett in 2001 was the beginning of the end. At the time it was concluded that 'Air New Zealand[']s ... decision to buy Ansett ... destroyed both'.²³ For today's financially troubled business, insolvency may be part of a considered business strategy. Craig Tractenberg identifies that '[f]ranchisors file for bankruptcy to escape or postpone the consequences of mass franchisee litigation, shareholder litigation, and lender enforcement activities'.²⁴ Other franchisor advisers concur with Tractenberg, acknowledging that voluntary administration can enable a franchisor

to reorganize its operations, deleverage its balance sheet, accomplish a sale of assets, obtain new financing or improve its capital structure. [It] may assist a franchisor in addressing ... overexpansion in the market and the need to eliminate units, an unworkable equity structure, desire to sell or merge with another entity, threat of franchisee litigation, [or a] desire to refinance [being hampered because] the lender has expressed concern about financial or other issues.²⁵

A franchisor's strategic insolvency²⁶ suggests a lack of fairness and good faith towards franchisees. It is an example of "capricious termination" that was identified

²⁰ John Stanworth, David Purdy and Stuart Price, 'Franchise Growth and Failure in the USA and the UK: A Troubled Dreamworld Revisited' (1997) 2(2) *Franchising Research: An International Journal* 75, 78–9.

²¹ James Cross, 'Franchising Failures: Definitional and Measurement Issues' (Paper presented at the International Society of Franchising Conference, Las Vegas, Nevada, 13–14 February 1994) 2–4.

²² Buchan, *Franchisees as Consumers*, above n 2, chs 7–8.

²³ Stathi Paxinos, 'Ansett from A to T', *The Sunday Age* (Sydney), 17 February 2002, 9.

²⁴ Craig R Tractenberg, 'What the Franchise Lawyer Needs to Know About Bankruptcy' (2000–01) 3(20) *Franchise Law Journal* 3.

²⁵ Sarah B Foster and Carolyn Johnsen, 'The War of the Worlds: Bankruptcy Versus ...' (Paper presented at the American Bar Association, 28th Annual Forum on Franchising, Florida, 19–21 October 2005).

²⁶ David Noakes, 'Measuring the Impact of Strategic Insolvency on Employees' (2003) 11(12) *Insolvency Law Journal* 91, 92, quoting Peta Spender, 'Scenes from a Wharf: Containing the Morality of the Corporate Law' (Paper presented at the 9th Annual Corporate Law Teachers Association Conference, Monash University, February 1999) 21–9.

by Harold F Brown in 1973 as [the] Achilles heel of the entire franchising industry'.²⁷ The disconnect between consumer protection of franchisees under the *Competition and Consumer Act 2010* (Cth) ('CCA') and the insolvency regime under the *Corporations Act 2001* (Cth) ('CA') is demonstrated in Figure 1 and will be explored later in this article.

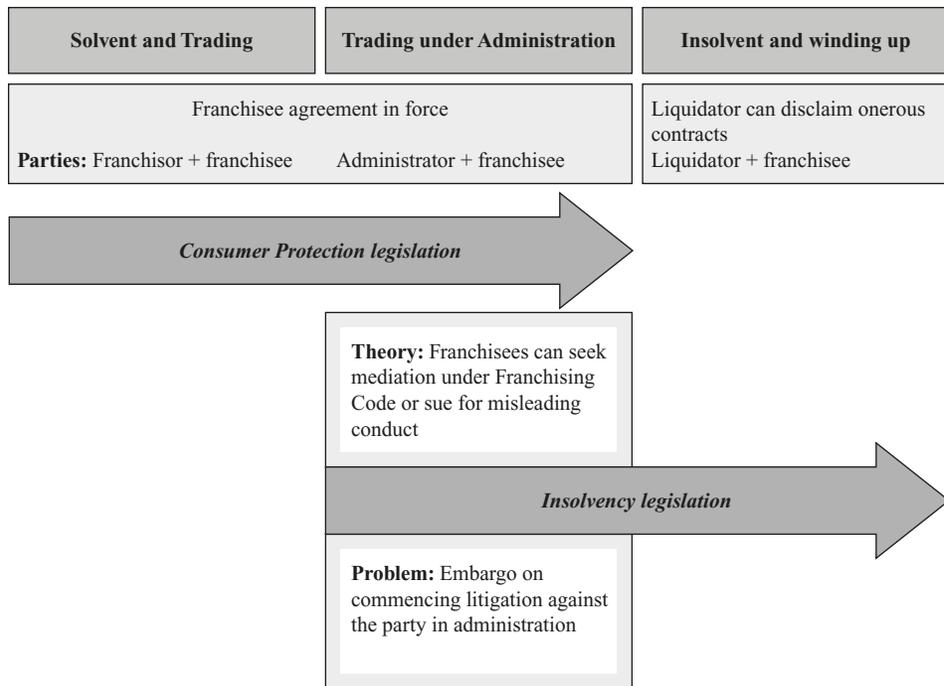


Figure 1: Insolvency legislation trumps contract, and consumer protection legislation²⁸

Australian franchise law still addresses the late 20th century franchise environment, but the reality of the franchising business model has progressed. As will be demonstrated by the A & R scenario, there are many stakeholders including public company and venture capitalist franchisor owners — and their shareholders and suppliers — whose conduct and legal rights may impact on the solvency of the franchise network and ultimately on the franchisee. Most are not factored into the *Trade Practices (Industry Codes — Franchising) Regulations 1998* ('Code'). This is acknowledged in the following comment:

²⁷ Harold Brown, *Franchising: Realities and Remedies* (Law Journal Press, 1973) 40, cited in Hunt, above n 8, 77, cited in Buchan, *Franchisees as Consumers*, above n 2, 123.

²⁸ Adapted from Jenny Buchan, *Franchisees as Consumers: Benchmarks, Perspectives and Consequences* (Springer, 2013) 151.

we've been talking about two parties in a relationship, the franchisor and franchisee, but we have really been talking about that in something of a vacuum. The reality for many of these operators is that there is a third player in the game — the Westfield or someone else — who has a relationship with one of those two, probably the franchisor. ... there is a risk-passing exercise which starts with the shopping centre owner, who then imposes a whole lot of costs on to the lessee of the premises, ... the franchisor. They then shove that down the line to the next person [the franchisee].²⁹

As well as risk-passing, financial problems such as cash flow or refinancing difficulties can be hidden elsewhere in the network of the corporate group, where they may incubate until they destroy the franchisor. In light of the above, an argument that the franchisees cause the failure of their franchisors is not sustainable. Accepting that franchisors can and do fail, it is useful to set out the recognised general and franchisor-specific indicators of insolvency.

III GENERIC AND FRANCHISOR-SPECIFIC INDICATORS OF A COMPANY'S INSOLVENCY

The cash flow test of solvency is adopted in the *CA* at s 95A. Under the cash flow test a debtor is insolvent when it cannot discharge all its debts when they fall due. Whether a debt has become due is not always determined by the time fixed in the contract; the date may depend also on industry practice.³⁰ The notion of 'become due' is a legal one and a debt is not rendered 'not due' just because the creditor has, to date, forborne from pursuing recovery.³¹ Before pursuing the A & R example, let us look at the indicators of both generic and franchisor-specific failure.

Australian courts have developed a generic checklist of indicators of corporate³² insolvency. In *ASIC v Plymin*,³³ a 'checklist of insolvency indicators was put in evidence'.³⁴ Justice Mandie used this checklist to determine whether the directors had breached the insolvent trading provisions of the *CA*.³⁵ The general indicators were: (1) continuing losses; (2) the liquidity ratio of the company being below 1; (3) overdue Commonwealth and State taxes; (4) poor relationship with banks, including inability to borrow further funds; (5) the company does not have access

²⁹ Economic and Finance Committee, above n 1, 80.

³⁰ *Re New World Alliance Pty Ltd (rec & mgr apptd); Sycotex Pty Ltd v Baseler [No 2]* (1994) 51 FCR 425, 434 (Gummow J).

³¹ *Melbase Corporation Pty Ltd v Segenhoe Ltd* (1995) 17 ACSR 187, 198–9.

³² In Australia 70 per cent of 'franchisor' entities are proprietary corporations, 14 per cent are public corporations and 10 per cent are trusts. Lorelle Frazer, Scott Weaven, Owen Wright, *Franchising Australia 2006* (Griffith University, 2006) 34.

³³ (2003) 175 FLR 124.

³⁴ Michael Murray and Jason Harris, *Keay's Insolvency Personal and Corporate Law and Practice* (Lawbook, 7th ed, 2011) 475.

³⁵ *Corporations Act 2001* (Cth) s 588G.

to alternative finance; (6) inability on the part of the company to raise further equity capital; (7) suppliers insisting on Cash on Delivery terms, or otherwise demanding special payments before resuming supply; (8) creditors being paid outside trading terms; (9) issuing of post-dated cheques; (10) the company has dishonoured some of the cheques paid to creditors; (11) special compromise arrangements made by the company with selected creditors; (12) solicitors' letters (of demands), summons(es), judgments or warrants issued against the company; (13) payments on the part of the company to creditors of rounded sums not reconcilable to specific invoices; and (14) inability to produce timely and accurate financial accounts to display the company's trading performance and financial position, and to make reliable forecasts.³⁶

While there is currently insufficient empirical data on the financial and risk profiles of failed (and successful) franchisors to provide a definitive list of indicators specific to franchising, 12 possible indicators of a franchisor's impending failure have been identified: (1) a large proportion of outlets being owned by the franchisor instead of franchisees, indicative of the return of failed franchisees to franchisor;³⁷ (2) a long history of failures on the part of franchisees in the franchise network;³⁸ (3) a breach of a franchisor's obligations to provide advertising support, equipment and inventory;³⁹ (4) evasiveness following franchisor default;⁴⁰ (5) a landlord's notice of demand;⁴¹ (6) a large number of court proceedings against the franchisor;⁴² (7) restructuring on the part of the franchisor, especially invoices from different

³⁶ (2003) 175 FLR 124, 213–14.

³⁷ Although Taylor and Hughes note that 'sometimes this is not apparent, as stores can be held by directors outside of the franchise structure': Fiona Taylor and Richard Hughes, *Five Lessons for Franchisees from the Kleenmaid Case* (6 October 2011) Griffith University <http://www.franchise.edu.au/articles/five-lessons-for-franchisees-from-the-kleenmaid-case.html?&utm_content=3846129>.

³⁸ Ray Borradale, *Why Some Franchisors Fail* (19 January 2009) Blue Maumau <www.blumaumau.org/why_some_franchisors_fail>.

³⁹ Craig R Colraine, *Franchises: Insolvency and Restructuring* (Paper presented at Ontario Bar Association CLE Distribution Law: Catch the Wave, Avoid the Rocks, Toronto, Canada, 26 May 2003) 3; Borradale, above n 38.

⁴⁰ Emily Maltby, 'Dragged into a Bankruptcy That Isn't Yours', *CNN Money* (online), 17 July 2009 <http://money.cnn.com/2009/07/17/smallbusiness/franchise_bankruptcy.smb>.

⁴¹ This may come direct to the franchisor as head tenant, and may not be passed on to the franchisee sub-tenant that is not in breach as it has paid rent to the franchisor.

⁴² 'Warning Signs of Failure', *The Australian Financial Review* (Sydney), 2 May 2006, 50. The significance of litigation as an early indicator of failure was identified in a Dun & Bradstreet Corporate Health Watch survey. It was concluded from the survey results that '[c]ompanies that have had legal action taken against them are nearly eight times more likely to fail than those that haven't'. Further, specific research would reveal whether the threat of franchisee litigation caused the franchisor to consider pre-emptive or strategic bankruptcy or whether the litigation caused the subsequent failure of the franchisor.

companies;⁴³ (8) franchisors not receiving previously favourable trading terms due to impending insolvency, especially where franchisees are ‘required to source stock or other services through their franchisor’;⁴⁴ (9) information in the franchisor’s balance sheet, the profit and loss statement, or announcements made to the stock exchange (where the franchisor is listed) pointing to an accumulation of significant debt when the franchise system is not expanding or the writing down of assets, or refinancing activities;⁴⁵ (10) information from credit reporting services about a franchisor company’s financial health;⁴⁶ (11) a failure on the part of the franchisor to make timely commission payments, where the franchisee is the franchisor’s commission agent; and as Canadian insolvency practitioner Craig R Colrairie suggests, (12) ‘[p]oor financial performance, including the accumulation of significant debt when the franchise system is not expanding, growing operating losses, the writing down of assets and re-financings’. Colrairie concedes that ‘[i]dentifying financial problems in non-publicly traded corporations⁴⁷ is more difficult’.⁴⁸ It can readily be deduced from the lists of general and franchise-specific indicators that many of the 26 are not apparent to franchisees.

Four factors differentiate franchisors from the corporations and corporate groups traditionally regulated under the *CA*. First, franchisors are able to raise additional equity capital through selling more franchises, and are thus able to hide their impending insolvency from franchisees for longer than they could hide it from a traditional finance source or from a supplier. Second, Abe de Jong, Tao Jiang and Patrick Wirwijmeren have found:

the maximum debt [to equity] ratio allowed [for franchisees in a system] depends on the size of the outlets, on the age of the franchise firm, on arrangements between the franchisor and the franchisee (such as cooperative advertising), and on the type of industry [and] that as the franchisor sets a higher requirement for the franchisee’s equity component than expected, the franchisor is [itself] able to raise more debt. ... the strategic use of the franchisee’s capital structure affects the franchisor’s decision of financing.⁴⁹

Third, there is an absence of direct scrutiny and accountability between franchisees and their franchisor because their respective governance structures are independent

⁴³ Taylor and Hughes, above n 37.

⁴⁴ Buchan, *Franchisor Failure*, above n 7, 48.

⁴⁵ Colrairie, above n 39, 3.

⁴⁶ For example, the Credit Reference Association of Australia, or Dunn and Bradstreet Inc. These sources should also be used with awareness that some of the information recorded will have been supplied by the franchisor itself.

⁴⁷ Best estimates indicate that in 2012 there were 40 franchisors owned by or trading as public companies in Australia with the remaining 1140 being proprietary companies, trusts, partnerships and sole traders, or combinations of these.

⁴⁸ Colrairie, above n 39, 3.

⁴⁹ Abe de Jong, Tao Jiang and Patrick Wirwijmeren, ‘Strategic Debt in Vertical Relations: Evidence from Franchising’ (2011) 87(3) *Journal of Retailing* 381–92.

of each other. Finally, franchisors owe no statutory duty of care and no fiduciary duty to their franchisees. Theirs is purely a contractual relationship. As a consequence of these differentiating factors, whilst any of the general indicators may be relevant in the case of a franchisor, they may not be apparent to its franchisees. For example, general indicator (10) was present in REDgroup, the ultimate owner of A & R, which ‘delayed payments to its suppliers and landlords by up to 30 days’.⁵⁰ Figure 3 identifies that in August 2010 REDgroup also breached financial covenants (concerning interest cover ratio and gearing ratio) and received a waiver from its lenders.⁵¹ A & R franchisees could not have learned of these breaches. Any of the 12 franchisor-specific indicators may have a benign explanation so should be treated cautiously.

The scarcity of highly reliable indicators available to franchisees is in direct contrast to the plethora of indicators of a franchisee’s impending insolvency that are available to a franchisor. The franchisor provides for itself a contractual right to receive numerous periodic payments from franchisees on time, to access franchisees’ figures and to conduct audits on all aspects of the franchisee’s business.

We will now turn to the demise of A & R.

IV ANGUS & ROBERTSON

The failure of franchisor A & R serves to illustrate the experience of franchisees whose failed franchisor becomes embedded in a complex corporate group. A & R’s franchisees were attracted to the honey pot and became firmly ensnared in the bear trap. A & R knew the business of buying and selling books.⁵² The company traded successfully as an Australian buyer and seller of books for 91 years before starting franchising, and over 100 years before its first merger in 1990. This longevity does not distinguish A & R from other failed franchisors.⁵³ As is demonstrated in Figure 3, A & R had several owners before 2004 when it was sold to a venture capitalist called Pacific Equity Partners (‘PEP’). The business structure within which franchising is conducted is often complex.⁵⁴ Through its involvement with PEP, for example,

⁵⁰ Ferrier Hodgson, ‘REDgroup Retail Pty Limited and Associated Companies (Administrators Appointed) Report by Administrators Pursuant to Section 439A(4)(a) of the *Corporations Act 2001* (Cth)’ (Administrators’ Report, 25 July 2011) 9.

⁵¹ *Ibid.*

⁵² Anne Wright, ‘Hundreds Sacked at Angus & Robertson’, *Herald Sun* (online), 16 June 2011 <<http://www.news.com.au/business/hundreds-sacked-at-angus-and-robertson/story-e6frfm1i-1226076157825>>.

⁵³ For example, Darrell Lea was placed in administration after trading for 85 years, Speeds Shoes after 94 years, Traveland after 43 years and Kleins Jewellery after 26 years.

⁵⁴ Similarly, failed franchisors Kleenmaid comprised two groups of companies with the franchisor entity nested on the periphery, and Traveland was one of over 40 entities under the Ansett umbrella.

Angus & Robertson Pty Ltd became one of a group of 22 companies and trusts, including two franchisors and one master franchisee, operating in three countries as shown in Figure 2.

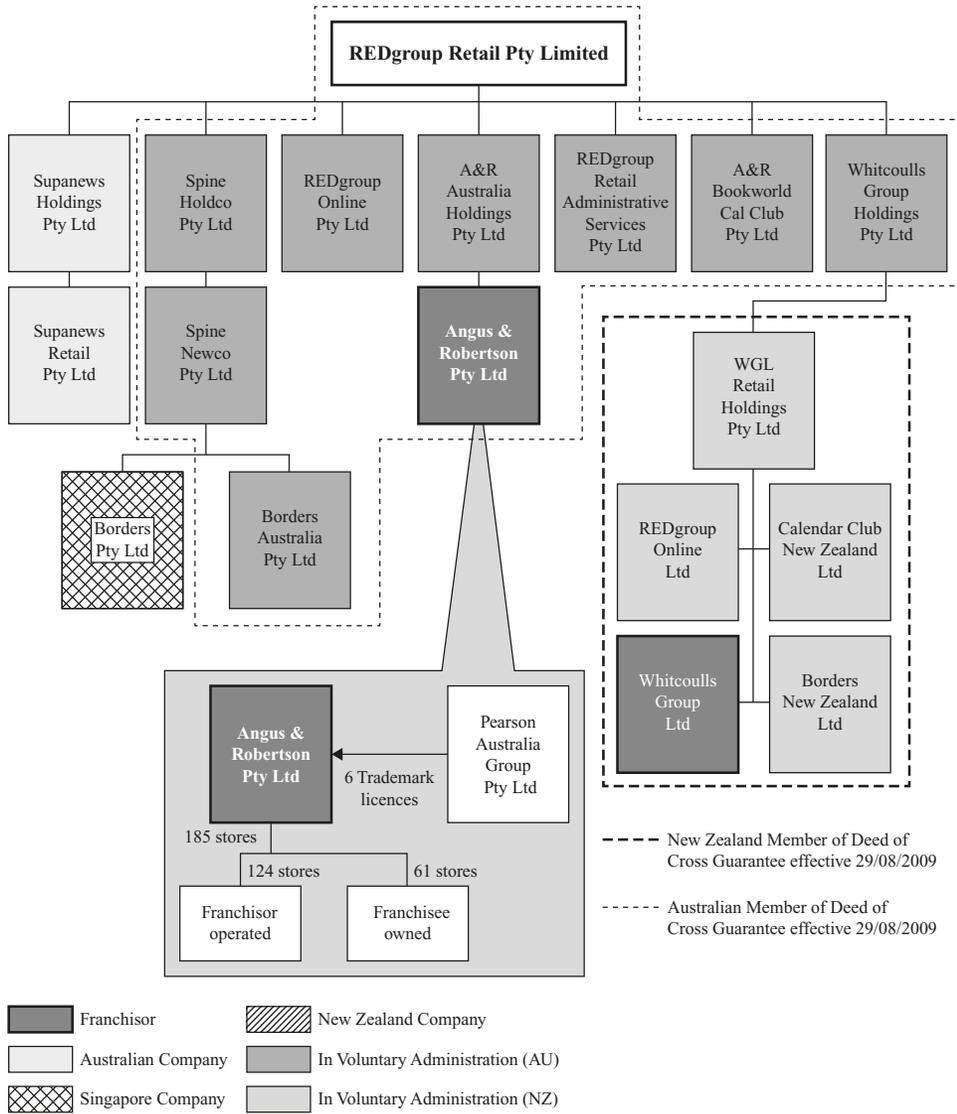


Figure 2: REDgroup organisation chart⁵⁵

Franchisees paid up to \$380 000 to establish an A & R store. This sum becomes ‘sunk’ costs. Franchisees also signed franchise agreements and premises leases.

⁵⁵ Adapted from Ferrier Hodgson, above n 50, app C (Group Structure), published in Jenny Buchan, *Franchisees as Consumers: Benchmarks, Perspectives and Consequences* (Springer, 2013) 122.

A & R franchise agreements provide a five-year term with a five-year renewal. Mr and Mrs Appleby purchased two A & R franchised bookstores in Queensland in November 2010. Stock is the franchisee's second major cash investment and costs \$900–1000 per square metre.⁵⁶ The Applebys' timing meant they were able to start strongly as

[t]he book industry ... relies heavily on the Christmas trade. [A] large percentage of our sales, and our positive cash flow comes from the Christmas season. Franchisees must be prepared for the increase in stock required to meet Christmas demand.⁵⁷

This example demonstrates how

[t]he risks associated with entering the franchise arrangement from the franchisee's point of view are predominantly financial. The initial step of setting up a franchise requires substantial up-front investment that usually covers the franchise fee, equipment costs, set-up fees and other expenses.⁵⁸

Moving now to the A & R insolvency: on 17 February 2011, within three months of buying their stores, and only two hours after completing their franchisee induction at REDgroup Retail network's head office in Melbourne, the Applebys received a call from their bank manager that their franchisor had appointed an administrator.⁵⁹ As new franchisees, the Applebys 'had done a tour of the office [the same day the administrator was appointed], [the REDgroup] had welcomed us, the CEO had been in to talk to us the day before. Mrs Appleby said "call me naive, but I can't believe that any of the staff in that building on that day knew"'.⁶⁰ Their experience highlights an important aspect of insolvency: it brews slowly and then happens fast. The franchisees can be the 'last to find out about a franchisor's financial problems'.⁶¹

A Why did A & R Fail?

Figure 3 shows the stages of growth and ultimate decline of A & R.

During 2010 it became clear to those in the sector that the environment for retailing books was becoming difficult. The well-known US-based book retail franchisor, Borders, filed for Ch 11 bankruptcy protection in the US in February 2011. Within

⁵⁶ A & R, *Terms of Agreement and Financial Commitment* <<http://www.angusrobertson.com.au/franchise-terms-of-agreement>>.

⁵⁷ A & R, *Being a Franchise Owner* <<http://www.angusrobertson.com.au/franchise-partners>>.

⁵⁸ Economic and Finance Committee, above n 1, 15.

⁵⁹ 'Light at End of the Tunnel: Good Staff and Family Support See Stores Write Another Chapter', *Sunshine Coast Daily* (Maroochydore), 13 September 2011, 20.

⁶⁰ *Ibid.*

⁶¹ Frank Zumbo, Submission to Parliament of South Australia, House of Assembly, Economic and Finance Committee, *Inquiry Into Franchises*, February 2007, 9.

24 hours, 'REDgroup Retail [(the Australian master franchisee of Borders)] was placed in administration ... owing an estimated \$AUD170 million'.⁶²

REDgroup had taken out a loan from PEP Fund IV, LP for \$138 million to complete the acquisition of Borders. 'The debt due to PEP was cross-collateralized across the group. REDgroup's secured creditors, ... PEP lodged a Proof of Debt in the Administration of each company for \$118 547 419'.⁶³ This enabled PEP as a secured creditor to appoint voluntary administrators. It should be noted that 'some franchisees would have entered the A & R system before PEP became the owner or before PEP created the REDgroup. Those that knew of the involvement of PEP might have been reassured to read that PEP described'⁶⁴ itself as 'a leading Australasian private equity firm focusing on buyouts and late stage expansion capital in Australia and New Zealand' that assists businesses move closer to their 'full potential'.⁶⁵ There were eight PEP entities.⁶⁶ Even if it had been aware of PEP's involvement it is beyond the resources of individual prospective franchisees to conduct due diligence on a set of complex entities. Additionally, the administrator's conclusion that there was no evidence of the REDgroup having traded whilst insolvent suggests that even the most diligent A & R franchisee would have found nothing to cause alarm.

REDgroup's directors attributed the failure to external events: consumer spending patterns, the GST and parallel importation laws giving online sellers an unfair benefit, and the strong Australian dollar 'which had appreciated against the US dollar and the pound sterling by 20% since September 2009'⁶⁷ and given Australians strong overseas buying power. The administrators added internal factors that contributed to the failure of the REDgroup and its franchise subsidiaries. These were

more emphasis on 'buying' than 'selecting' stock resulting in overstocking with aged, poor stock; failure to recognize and promptly address loss making stores; under-utilisation of space in stores and poor organisation with no logical grouping

⁶² Jenny Buchan, 'The Failure of Pre-Purchase Disclosure to Protect Franchisees of a Franchisor in Administration' in Thomas Ehrmann et al (eds), *Network Governance: Alliances, Cooperatives and Franchise Chains* (Springer-Verlag, 2013) 313, 317.

⁶³ Ferrier Hodgson, above n 50, 7.

⁶⁴ Buchan, 'The Failure of Pre-Purchase Disclosure', above n 62, 325.

⁶⁵ Pacific Equity Partners, *About Us* <<https://investor.pep.com.au/pages/content.asp?pid=1>>; cf Pacific Equity Partners, *About Us* <<http://www.pep.com.au/about-us/>>.

⁶⁶ Pacific Equity Partners Fund I Pty Limited ACN 083 026 822, Pacific Equity Partners Pty Limited ACN 082 283 949, Pacific Equity Partners Fund III (Australasia) Pty Limited ACN 117 565 410, Pacific Equity Partners Fund II (Australasia) Pty Limited ACN106 318 370, Pacific Equity Partners Fund 111 GP (Jersey) Limited 126745686, Pacific Equity Partners Fund III SPV Pty Limited ACN 119 059 040, Pacific Equity Partners Fund IV (Australasia) Pty Limited ACN 124 839 989 and Pacific Equity Partners Fund IV LP 150258165.

⁶⁷ Ferrier Hodgson, above n 50, 13–14.

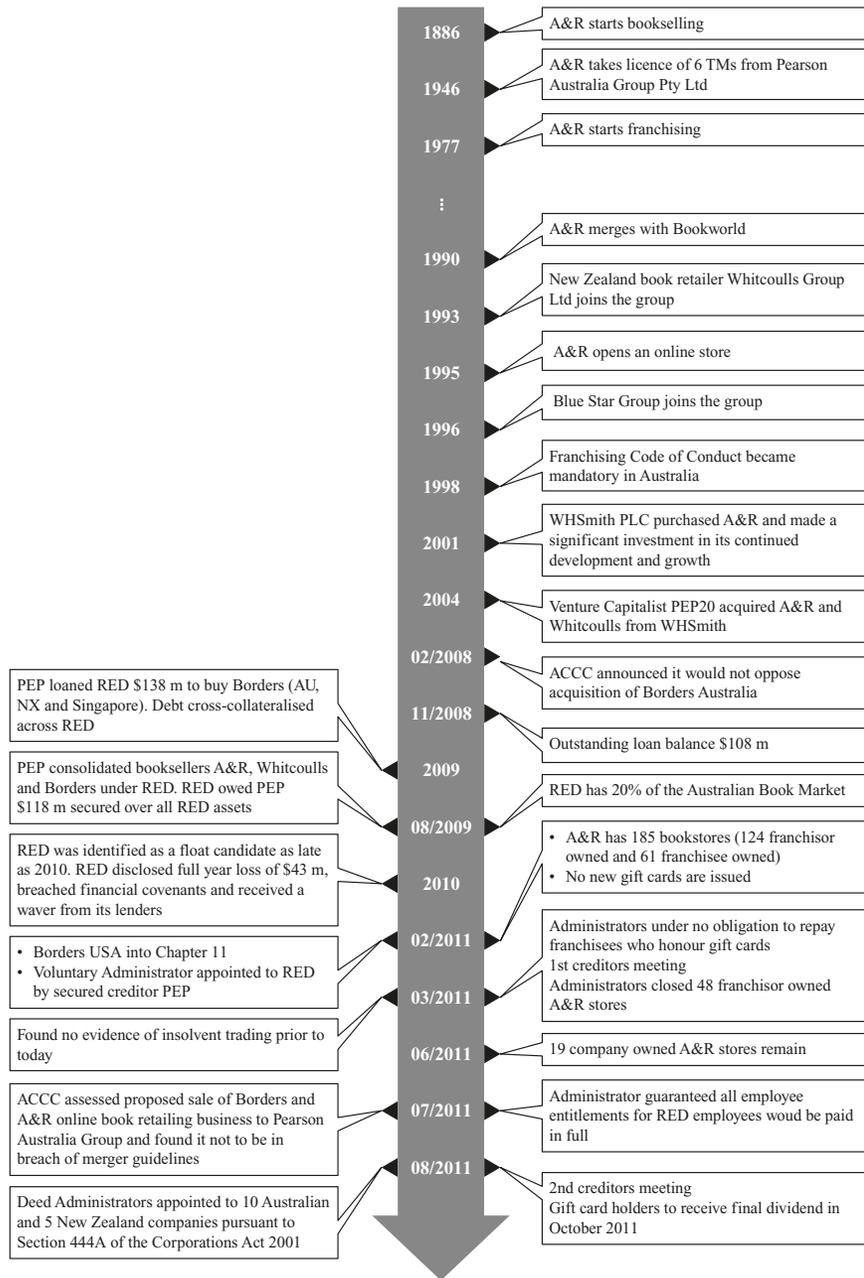


Figure 3: Rise and fall of Angus and Robertson⁶⁸

⁶⁸ From Jenny Buchan, *Franchisees as Consumers: Benchmarks, Perspectives and Consequences* (Springer, 2013) 126. The references in the timeline in Figure 3 are from the following sources: references dated 1886, 1977, 1990, 1993, 1995, 1996, 2001, 2004: A & R, *About* <<http://www.angusrobertson.com.au/company-information>>; 1946: see trade mark registrations numbered 299489 (Class 16), 343650 (Class 16), 637633 (Class 42), 861016 (Class 16, 35, 41), 1025323 (Class 16, 35, 41) and 1073382

and general lack of consistent business processes with little use and reference to signed off critical paths and event management cycles.⁶⁹

As already mentioned, insolvency may be a strategic business decision. A New Zealand journalist, commenting on the impact of REDgroup's failure on the Whitcoulls franchise, suggested that PEP, upon realising that the planned exit strategy of floating had become unviable, saw voluntary administration:

as a cost-effective way ... to exit its ill-fated foray into book retailing. They ... used the law to the maximum possible extent to extract everything they could out of it. ... Effectively this was a staged exit. ... REDgroup's total secured debt was \$A118 million (\$NZ147 million), most of it owed to [secured creditor] Pacific Equity Partners.⁷⁰

Voluntary administration meant REDgroup could walk away from the franchisees without having to buy businesses back or risk being sued. The likelihood that this voluntary administration was a strategic move is given weight by the REDgroup administrators' conclusion that there was no evidence of trading while insolvent. The

(Class 16, 35, 41); 02/2008: Australian Competition & Consumer Commission, 'A&R Whitcoulls Group Holdings Pty Ltd (ARW) — Proposed Acquisition of Borders Australia Pty Limited' (Public Competition Assessment, 27 February 2008) 1; 11/2008, 2009: adapted from Ferrier Hodgson, above n 49, 6; 08/2009: Ferrier Hodgson, above n 49, 7; 08/2009: Chris Zappone, 'Borders, A&R Call Time on Book Vouchers', *The Sydney Morning Herald* (Sydney), 23 March 2011; 2010: Ferrier Hodgson, above n 49, 9; 02/2010: Patrick Stafford, 'Borders Australia and Angus & Robertson Chains Collapse', *Smartcompany* (online), 17 February 2011 <<http://www.smartcompany.com.au/growth/economy/18695-20110217-borders-australia-and-angus-robertson-chains-collapse.html>>; 02/2011: Ferrier Hodgson, 'A&R Has 185 Bookstores', 'No New Gift Cards Are Issued', *Frequently Asked Questions: Franchisees* (17 February 2011); 03/2011: Ferrier Hodgson, 'Second Meeting of REDgroup Creditors' (Press Release, 27 July 2011); 03/2011: Ferrier Hodgson, 'Administrators Under No Obligation to Repay', '1st creditors meeting', 'Administrators Closed 48 Franchisor Operated A&R Stores', *Frequently Asked Questions: Franchisees* (17 February 2011); Chris Zappone, 'Booksellers' Woes Worsen as Franchisees Seek to Defect', *The Sydney Morning Herald* (Sydney), 5 April 2011; 06/2011: Chris Zappone, 'Pearson Buys Borders, A&R Websites', *The Sydney Morning Herald* (Sydney), 5 July 2011; 07/2011: ACCC, *Mergers Register: Pearson Australia Group Pty Ltd — Completed Acquisition of Certain Assets of REDGroup Retail Pty Ltd (Administrators Appointed)* (August 2011) <<http://www.accc.gov.au/content/index.phtml/itemId/1002218/fromItemId/751043>>; 07/2011: Ferrier Hodgson, above n 49, 19; 08/2011: see DOCA dated 4 August 2011 between REDgroup Retail Pty Limited (Administrators Appointed) and the companies listed in Part A of Schedule 3 and Steven John Sherman, John Melluish and John Lindholm of Ferrier Hodgson; 08/2011: Ferrier Hodgson, 'Angus & Robertson Pty Limited and Borders Australia Pty Limited (Subject to Deed of Company Arrangement) ('the Companies'), Gift Cards', *Frequently Asked Questions 2*; Ferrier Hodgson, above n 49, 29.

⁶⁹ Ibid.

⁷⁰ Jamie Gray, 'REDgroup Creditors Set To Vote On Slim Payout', *The New Zealand Herald* (Auckland), 4 August 2011.

administration was voluntary, triggered by a secured creditor PEP that owned the majority of shares in the REDgroup. The consequential A & R administration can arguably be categorised as part of a considered business strategy. Voluntary administration provided an opportunity for the venture capitalist PEP to exit its investment at relatively low cost. This failure could not be attributed to the franchisees' conduct. Faced with a failing franchisor, what can franchisees do to protect their businesses?

V WHAT HAPPENS WHEN A FRANCHISOR FAILS?

On failure of the franchisor, the group arrangements where economies of scale can be achieved cease. The cost of doing business will accordingly increase for franchisees that become standalone operators. The services previously provided by franchisors will have to be taken in-house.⁷¹ In some cases relationships will have to be forged with new suppliers who may be reluctant to supply to an unproven operator. What role does the law play?

A Contracts

The franchise agreement is the primary regulatory source of the franchise relationship.⁷² The franchisees also assume, wrongly, that there will be statutory protection through the *Code* and the *CA*. Each source of regulation is now examined.

A franchisor is 'connected' to its franchisees by contracts, including a franchise agreement. The party (franchisor) drafting a standard form commercial contract acts in its own interests. The franchisee of a single unit is offered the opportunity to buy into the franchise on a 'take it or leave it' basis. A franchisor's contractual duties are typically expressed as discretions and the franchisees' duties are expressed as obligations with clear consequences for breach. The franchise agreement is only one of many contracts franchisees enter into in order to operate their business. The absence of balance in the initial contract is symptomatic of deep asymmetry that permeates franchise relationships.⁷³ Franchisees also sign leases, sub-leases or premises licenses, supplier agreements, licences granting them the right to use intellectual property and employment contracts with their own staff. These consequential contracts where, again, the franchisee is typically not the party drafting the contract, seldom if ever identify franchisor failure as a trigger that would enable the franchisees to terminate without penalty. Franchisees must continue to perform all contracts even if the lynch pin, the franchisor, has failed and is unable to support them. For example, as illustrated by this comment by a former franchisee:

⁷¹ Birkett Long Solicitors, above n 15.

⁷² Elizabeth Crawford Spencer, *The Regulation of Franchising in the New Global Economy* (Edward Elgar, 2010) 76–115.

⁷³ Jenny Buchan 'Ex Ante Information and Ex Post Reality for Franchisees: The Case of Franchisor Failure' (2008) 36(6) *Australian Business Law Review*, 407–31 explores contract, process, information, regulatory and risk asymmetry in franchising.

[w]e had just signed renewals of our franchise agreements with Traveland (this franchisee had multiple agencies) for another 5 years when Ansett became insolvent. ... There was no way we could get out of our franchise agreements or premises leases.⁷⁴

Contracts, including the franchise agreement, are not terminated by the appointment of an administrator. Their position under leases is very important for affected franchisees. There are numerous models through which a franchisor secures retail premises and then provides tenure to franchisees. These include the franchisor owning the premises or taking the head lease; the franchisee taking the head lease, a sub-lease or a licence; or the franchisee owning the premises.⁷⁵ In some situations the franchisee guarantees the performance of the franchisor under the head lease.⁷⁶ Leases 'present [an] area of recurring uncertainty'⁷⁷ to administrators, to whom

[s]ection 443B(2) of the CA⁷⁸ [provides] a grace period of five business days before [the administrator] incurs personal liability for rent under pre-appointment leases. During [that time] the administrator can give the lessor a notice that the company does not propose to exercise rights in respect of the property (s 443B(3) CA).⁷⁹

A notice served on a landlord under s 443B(3) 'does not terminate the lease'.⁸⁰ A further complication for franchisees is that where the franchisee sub-tenant has guaranteed the performance of its franchisor under a head lease, and the franchisor has not paid the landlord, the franchisor's default makes the franchisee guarantor liable. This can mean the franchisee pays the same rent twice: once to the franchisor, and again as guarantor. Regardless of how the administrator treats the lease, payments due by the franchisee to the franchisor continue to be payable, as do payments by the franchisee to third parties. If the franchisor is wound up, the liquidator has the power to disclaim⁸¹ a lease as an onerous contract. This leaves the franchisee without tenure.

⁷⁴ Jenny Buchan and Lorelle Frazer, 'The Domino Effect: How Ansett Airlines' Failure Impacted on Traveland Franchisees', (Paper presented at the Academy of World Marketing Management and Development Conference, Paris, 10–13 July 2006) 1907.

⁷⁵ See Jenny Buchan and Bill Butcher, 'Premises Occupancy Models for Franchised Retail Businesses in Australia: Factors for Consideration' (2009) 17(2) *Australian Property Law Journal* 143, 170 for a discussion of the combinations and some consequences.

⁷⁶ See, eg, *Neldue Pty Ltd v Moran* [2004] WASC 100 (8 April 2004); *Miller v Loyal [No 6]* (2001) 187 ALR 766.

⁷⁷ Orla M McCoy, 'Administrators and Leases: Obligations and Options' (2012) 24(2) *Australian Insolvency Journal* 24.

⁷⁸ CA div 9: Administrator's liability and indemnity for debts of administration.

⁷⁹ McCoy, above n 77, 25.

⁸⁰ *Silvia v Fea Carbon Pty Ltd (admin apptd) (recs and mgrs apptd)* (2010) 185 FCR 301.

⁸¹ CA s 568; see *Willmott Growers Group Inc v Willmott Forests Ltd (in liq) (recs and mgrs apptd)* [2013] HCA 51 (4 December 2013).

Franchisees may also be affected by retention of title arrangements. Often franchisees believe they have title to stock on site, and have included it in calculations of assets available as security to their financiers. However, franchise agreements may stipulate that stock is not transferred to the franchisee until it is paid for.

B *The Code*

In singling out franchisees for protection from exploitation by franchisors policy makers have recognised that unlike suppliers that are able to negotiate their terms of trade, or employees who have extensive statutory protections, or consumers who can return a faulty product, franchises are uniquely vulnerable.⁸² Having signed a franchise agreement, a franchisee submits to its franchisor's significant power and discretion. The *Code* purports to protect franchisees in key risk areas, both ex ante and ex post signing a franchise agreement.

Compliance with the *Code* requires a franchisor to attempt to resolve disputes via mediation, and to provide a prospective franchisee with pre-contract disclosure. Requests by franchisees for mediation during the administration are not well received by administrators.⁸³ The *Code* enables a franchisor to instantly terminate the franchise agreement if a franchisee becomes bankrupt, insolvent under administration or an externally administered body corporate,⁸⁴ but the converse does not apply. This lopsided statutory right entrenches the asymmetrical nature of the relationship. Only rarely is a counterbalancing *ipso facto* clause found in a franchise agreement to provide the franchisee with the right to terminate if its franchisor exhibits signs of financial ill health.⁸⁵

The *Code*-prescribed disclosure is current information, predominantly addressing the financial and legal fitness of the entity identified as 'the franchisor'. It is difficult for a franchisee to objectively verify at a reasonable cost much of the information disclosed by the franchisor, so franchisees rely heavily on the information supplied in the disclosure document. To return to A & R, its disclosure would have provided franchisees with a plethora of information under 23 major headings. Since 2010 the front cover of the disclosure made by franchisors warns:

⁸² See above nn 39–66; Buchan, *Franchisees as Consumers*, above n 2; Jenny Buchan, 'Square Pegs in Round Holes: Franchisees of Insolvent Franchisors' (2008) 2(9) *Business Law International* 114, 121–2.

⁸³ Their reasons and the opposing arguments are outlined in Buchan, *Franchisor Failure*, above n 7, 207–8.

⁸⁴ A franchisor does not have to comply with cl 21 or 22 [ie provide notice of intention to terminate] if the franchisee: ... (b) becomes bankrupt, insolvent under administration or an externally administered body corporate': *Code* cl 23 (Termination — special circumstances).

⁸⁵ Jenny Buchan and Bill Butcher, 'Premises Occupancy Models for Franchised Retail Businesses in Australia: Factors for Consideration' (2009) 17(2) *Australian Property Law Journal* 143, 170. From a sample of 70 franchise agreements on USA database Free Franchise Docs <<http://www.freefranchisedocs.com/index.html>> as at 5 June 2008, 76 per cent permitted the franchisor to terminate the franchise agreement, usually with no notice and no right to cure, if the franchisee became bankrupt.

Entering into a franchise agreement is a serious undertaking. Franchising is a business and, like any business, the franchise (or franchisor) could fail during the franchise term. This could have consequences for the franchisee.⁸⁶

This is unlikely to promote caution in franchisees buying into a long-established franchisor. The specific disclosure items with the potential to reveal information about the franchisor, its plans and its attitude to risk are Items 2 (franchisor's details), 3 (franchisors' business experience), 4 (litigation), 7 (intellectual property) and 18 (obligation to sign related agreements).

'Franchisor' and 'Associate' are defined terms under Item 2. The Appleby's franchisor was A & R. Under the heading of associate the franchisor must disclose a person

- (a) who:
 - (i) is a director or related body corporate, or a director of a related body corporate, of the franchisor; or
 - (ii) for a franchisor that is a proprietary company — directly or indirectly owns, controls, or holds with power to vote, at least 15% of the issued voting shares in the franchisor; or
 - (iii) is a partner of the franchisor; and
- (b) whose relationship with the franchisor is relevant to the franchise system, including supplying goods, real property or services to a franchisee.⁸⁷

Without access to the A & R disclosure document it is not clear whether the existence of controlling shareholder PEP would have been disclosed. PEP would not have been supplying goods, real property or services to the franchisees. It can be seen from Red Group's organisational chart (Figure 2) that PEP was probably too distant to be disclosed. It had a role superficially unrelated to supporting the franchisees. The franchisees could have had no knowledge of the \$AUD138 million debt, or the manner in which it was secured. If all franchise agreements were included in this security then this diminishes their value as security for loans taken out by franchisees to purchase their businesses. Even if they had known of the debt to PEP, franchisees would not have the ability or the resources to evaluate its significance. Any disclosure provided before 2009 would have preceded the existence of the REDgroup. Due diligence is further discussed under the heading 'Franchisees' Ex ante and Ex post Responses to Franchisor Insolvency'.

Item 3 requires the franchisor to outline its business experience. As previously noted, A & R was long established. It has already been franchising for 33 years when the Applebys signed franchise agreements. The parent, REDgroup, had the appearance of being a very well organised, well-capitalised, geographically diversified player in the retail bookselling world.

⁸⁶ Code Annexure 2 item 1.1(e).

⁸⁷ Ibid cl 3(1).

Item 4 addresses litigation. It would not have revealed any current litigation against the franchisor or its associates. The *Code* directs franchisors and franchisees to attempt to resolve disputes through mediation. As disputes resolved by mediation are confidential, and mediation is not litigation, their existence is not disclosed.⁸⁸ A breach of a contract with a third party⁸⁹ may be a ‘red flag’ about impending insolvency, but it does not need to be disclosed unless it is currently being litigated. Item 4 would not give rise to disclosure of a franchisor’s debt unless that debt has triggered litigation. For A & R franchisees, therefore, Item 4 would not have drawn attention to REDgroup’s breaches of financial covenants, or its receipt of a waiver from its lenders in 2010. These events approached general insolvency indicator (4), but remained hidden from franchisees.

Item 4.1(a)(iii) requires any franchisor contravention of the *CA* to be disclosed. Trading while insolvent is a contravention of the *CA* but A & R was not thought by the administrator to have been trading while insolvent.

Under Item 7 (intellectual property), A & R would have revealed the existence of six registered trademarks⁹⁰ protecting the names ‘Angus & Robertson’ and ‘Angus & Robertson where books come to life’, all owned by a company called Pearson Australia Group Pty Ltd (‘Pearson’) and licensed to A & R. The appointment of the administrators to the licensee is likely to have been an event allowing Pearson to terminate the trademark licenses. Pearson ultimately bought the REDgroup’s online retailing business from the administrators after gaining mergers approval from the Australian Competition and Consumer Commission (‘ACCC’).⁹¹

The ‘[o]bligation to sign related agreements’ identified in Item 18 would direct the franchisees’ attention to its premises lease. In some franchise systems the franchisor might also require franchisees to lease fit out or to sign loan agreements and personal guarantees.

Item 2 of the *Code* provides for a pre-purchase statement about the financial details that have been supplied to the franchisee. A director of the franchisor is required

⁸⁸ For a discussion of mediation and the problems the confidentiality of the process creates see Jenny Buchan, Jennifer Harris and Gehan Gunasekara, ‘Franchise Mediation: Confidentiality or Disclosure: A Consumer Protection Conundrum’ (Paper presented at the 25th Annual International Society of Franchising Conference, Boston, USA, June 2011).

⁸⁹ See Patrick Stafford, ‘Borders Australia and Angus & Robertson Chains Collapse’ (17 February 2011) Smart Company <<http://www.smartcompany.com.au/growth/economy/18695-20110217-borders-australia-and-angus-robertson-chains-collapse.html>>.

⁹⁰ Numbered 299489 (Class 16), 343650 (Class 16), 637633 (Class 42), 861016 (Class 16, 35, 41), 1025323 (Class 16, 35, 41) and 1073382 (Class 16, 35, 41).

⁹¹ ACCC, *Mergers Register: Pearson Australia Group Pty Ltd — Completed Acquisition of Certain Assets of REDGroup Retail Pty Ltd (Administrators Appointed)* (August 2011) <<http://www.accc.gov.au/content/index.phtml/itemId/1002218/fromItemId/751043>>.

to sign the statement that in the director's opinion the franchisor is solvent. From franchisors that are not trading strongly this may be of limited value. Usually, only public corporations in Australia must be audited. Even if the auditor has identified a situation that casts doubt on an entity's 'going concern' status, the directors may have been able to satisfy the auditor that there are mitigating circumstances and that 'all will be well'. Such mitigating circumstances, for instance new franchisees committed to investing, may or may not eventuate. Thus, a *Code*-compliant audit may present an inaccurate and ultimately misleading picture of the franchisor's solvency.

The *Code*, with its emphasis on 'the franchisor' remains a response to franchising as it was modeled in the 20th century. Figure 2 demonstrates how nuanced the franchising model has become.

C Corporations Act

A transition from consumer protection under contract law and the *Code* to insolvency administration and thence to winding up under the *CA* signal a shift in focus, and in statutory duties. Importantly for franchisee advisers, whilst administrators are bound by the *Code*, and the terms of the franchise agreement,⁹² they are regulated by the *CA*. The liquidators' duties and liabilities are found only in the *CA*.⁹³

The *CA* provides three possible paths through the insolvency process.⁹⁴ The most common corporate insolvency procedure is voluntary administration resulting in a deed of company arrangement ('DOCA') or liquidation. Voluntary administration is an external administration where the directors of a financially troubled company or a secured creditor with a charge over most of the company's assets appoint an external administrator, a 'voluntary administrator'. The voluntary administrator must investigate the company's affairs, and report to creditors, recommending whether the company should enter into a DOCA, go into liquidation or be returned to the directors.

A voluntary administrator is usually appointed by a company's directors after they decide that the company is insolvent or likely to become insolvent. Less commonly, a voluntary administrator may be appointed by a liquidator, provisional liquidator, or a secured creditor. Once the administrator has been appointed, factors that impact on how individual franchisees fare include the financial climate at the time of the failure, the proximity of competitors, the type of product or service the franchisee's business sells, whether the franchisor's business is sold to one buyer or the assets are

⁹² See *Schering Pty Ltd v Forrest Pharmaceutical Co Pty Ltd* [1982] 1 NSWLR 286, where the court granted an ex parte injunction restraining a proposed breach of a franchise agreement by a franchisor in receivership. See also Steve McAuley, 'Squeezing the Lemon Dry' (2010) 48(3) *Law Society Journal* 61.

⁹³ See Michelle Power, Nick Anson and David Dickens 'Liquidators of Landlords Can Disclaim Tenants' Leasehold Interests' (2012) 50(10) *Law Society Journal* 60, 60–3.

⁹⁴ See Australian Securities & Investments Commission, *Resources* <<http://www.asic.gov.au/asic/asic.nsf/byheadline/Resources>>.

sold to several unrelated buyers, and the expertise of the buyer of the franchisor. As one insolvency practitioner remarked:

[s]ale of the [franchisor's] business may subject the franchisee to the control of a company unfamiliar in the area and incapable of running the business profitably. The dramatic demise of Traveland [franchise subsidiary of Australia's former Ansett Airlines] demonstrates the implications of a buying entity that has little experience in the franchisor's core business area and has insufficient expertise or resources to support the business.⁹⁵

The DOCA is the most common outcome of administration. A DOCA is a binding arrangement between a company and its creditors, governing how the company's affairs will be dealt with. It aims to maximise the chances of the company — or as much as possible of its business — continuing, or to provide a better return for creditors than an immediate winding up of the company, or both.⁹⁶

If an administrator recommends that a business cannot be saved and should be wound up, control of the company passes to the liquidators. Liquidation is the orderly winding up of a company's affairs. The liquidator sells the company's assets, ceases or sells its operations, distributes the proceeds among the creditors and distributes any surplus to the shareholders.

Regardless of the specific path the insolvency follows, s 471B of the *CA* applies to deny any right to continue or initiate court action, except by the administrator or with the consent of the court.⁹⁷ The duties on the administrator or liquidator thus effectively trump the consumer protection regime (the *CCA*, *Code* and any industry code enacted under the *Small Business Commissioner Act 2011* (SA)). During the administration (see Figure 1) the administrator, standing in the shoes of the franchisor, theoretically should engage with the franchisees, but the rigorous statutory time limits that apply in the insolvency arena, the provisions of the *CA* that focus on creditors' rights, the fact that franchisees are often spread far and wide geographically, and the absence of funds weigh in favour of the insolvency regime taking priority over consultation in practice.

Franchisees have difficulty achieving standing in their franchisor's insolvency through the *CA* because of their limited creditor status. Whilst the *pari passu* principle states

⁹⁵ Wayne Jenvey, 'Rocky Roads and Rollercoasters — Turnaround Strategies for Distressed Franchise Systems' (Paper presented at the Legal Symposium at the Franchise Australia Annual Conference, Gold Coast, 24 October 2006) 9.

⁹⁶ Australian Securities & Investments Commission, *Resources*, above n 94. See also Australian Securities & Investments Commission, *Insolvency Information Sheets* <<http://www.asic.gov.au/asic/asic.nsf/byheadline/Resources++Insolvency+information+sheets>>.

⁹⁷ The relevant legislation is the *CA* ss 440D, 471(2). See *Ibbco Trading Pty Ltd v HIH Casualty & General Insurance Ltd (in prov liq)* (2001) 19 ACLC 1093. See also Christopher Symes and John Dunns, *Australian Insolvency Law* (LexisNexis Butterworths, 2009) 295.

‘creditors should share equally in the bankrupt estate’,⁹⁸ the bulk of the franchisee’s investment is in the sunk costs of establishing its business. For these sums the franchisee is not a creditor. Only creditors have standing in a creditors’ meeting. The approach some administrators take to franchisees is that, if in doubt, put each franchisee in as a creditor for \$1. While this secures franchisees the right to attend creditors’ meetings, it pays scant recognition to their investment and potential loss.

Putting aside their initial investment, franchisees may manifest disparate claims as creditors or debtors. The Kleins franchise agreement, for example, contained a guarantee that if annual turnover did not reach a certain level, the franchisor would pay the franchisee an agreed sum. Anecdotally, many franchisees received or were entitled to receive these payments. For this sum they were a creditor. The franchisor and franchisee might have concluded a mediation agreement that resulted in the franchisee being owed money, or, similarly, moneys may be owed under a judgment debt. Some of the Traveland franchisees were owed money by the franchisor in connection with airline ticket refunds. A franchisee may also be a creditor in relation to any rent payments that the franchisor received from the franchisee but did not pass on to the landlord, arguably held by the franchisor on trust.

Widespread media attention to a failure can damage franchisees whose businesses may still be viable. News of the failure or impending failure of a franchisor will affect the franchisees’ relationship with their customers. The extent of the impact will depend on variables such as the length of time between the usual purchase and delivery of the product or service sold by franchisees, the way the customer finds out about the failure and the extent of franchisees’ rights to continue operating in reliance on their contracts with third parties. A franchisee selling white goods or travel⁹⁹ that will be paid for now and delivered in the future, for instance, will struggle to retain their customers’ confidence, whereas the franchisees selling cups of coffee that are consumed immediately should be able to continue trading so long as their suppliers continue supplying coffee beans. In the worst-case scenario, a franchisee’s customers will only discover the franchisor has failed when the franchisor’s head lease is disclaimed by the liquidator with the result that the franchisee is evicted from its premises and the customers are met by a locked roller door.

Under s 568(1) of the *CA*, liquidators (but not administrators) have the power to disclaim onerous contracts.¹⁰⁰ This enables them to disclaim head leases, trademark licences, franchise agreements, obligations to pay commissions and any other contracts that are seen as unsaleable or a drain on resources. Decisions the administrator or liquidator takes in relation to disclaiming onerous contracts will impact significantly

⁹⁸ See *ibid* 150–67 for an explanation of the *pari passu* principle and its exceptions.

⁹⁹ Buchan and Frazer, above n 74, 1908.

¹⁰⁰ It should be noted that the options in relation to leases and other onerous contracts available to administrators and liquidators differ. Administrators operate under pt 5.3A of the *CA* s 443A (B) and do not have a power that liquidators have under s 568 to disclaim onerous contracts. For a detailed discussion of this point see McCoy, above n 77, 24–9.

on the options available to franchisees. If the liquidator disclaims the head lease, the sub-lessee franchisee will lose the value of the sunk costs unless it is able to negotiate a new lease. Even then, the franchisee may find that the lack of the power of the brand or the loss of group bargaining power may render its business unviable.

If the administrator decides to wind up the franchisor's business, there is nothing to stop a liquidator selling the franchisor's business to a direct competitor¹⁰¹ of the franchisor. It is unlikely that an acquisition would meet the threshold test of, 'having the effect, or be likely to have the effect of substantially lessening competition in a market',¹⁰² that would lead to close examination of a proposed merger of two franchise networks by the ACCC. That direct competitor may elect not to buy the franchise agreements but, instead, to simply buy the brand cheaply and shelve it.

D The A & R Franchisees

A & R franchisees expected to be able to amortise their investment over a 10 year term but discovered '[t]he majority of those costs cannot be fully recovered in the case of the franchise's failure'.¹⁰³ Franchisees reported experiencing 'serious falls in sales after the collapse'.¹⁰⁴

A & R's administrators sent franchisees a circular early in the administration period stating that the appointment of the administrators did not automatically terminate the franchise agreements and royalties should continue to be paid by direct debit from franchisees' banks, as usual. The absence of strict duties on the franchisor enables administrators to require the ongoing performance of franchise agreements by franchisees so long as the franchisor is not in breach of a head lease or other essential supply line contract. From the administrators' perspective, franchisees 'continue[d] to trade as normal'¹⁰⁵ during the administration.

Where franchisees are selling instant use items like books, the administrator can be assured of a steady stream of revenue by requiring franchisees to keep trading. This was the situation in A & R's case. Before it was placed into voluntary administration the REDgroup network had over 2500 employees.¹⁰⁶ Despite the lack of ongoing support by the franchisor, post-administration royalties and marketing

¹⁰¹ CCA s 50 prohibits acquisitions that would result in a lessening of competition but the tests are relaxed where the entity being acquired is insolvent.

¹⁰² CCA s 50(1).

¹⁰³ Economic and Finance Committee, above n 1, 15.

¹⁰⁴ Madeleine Heffernan, '25 Angus & Robertson Franchisees To Go Independent After Collapse' (5 April 2011) Smart Company <<http://www.smartcompany.com.au/growth/franchising/19414-20110405-25-angus-robertson-franchisees-to-go-independent-after-collapse.html>>.

¹⁰⁵ Ferrier Hodgson, 'Negotiations Underway for Sale of Angus & Robertson Bookstores' (Press Release, 15 June 2011).

¹⁰⁶ 'Troubled Bookstores Face Closure This Week', *ABC Melbourne* (online), 1 March 2011 <<http://www.abc.net.au/news/stories/2011/03/01/3151898.htm>>.

contributions paid by franchisees contributed \$226 518 to the administrators' pool of funds. Ultimately 'stock realisations [being commissions on post-administration sales paid by franchisees were] sufficient to pay in full all employee entitlements, totaling approximately \$11.7 million'.¹⁰⁷

Within a month of their appointment the administrators closed '48 franchisor-owned A & R stores'.¹⁰⁸ They identified the A & R franchise agreements as saleable assets. The franchisees were required to continue trading whilst the administrators sought a buyer for the businesses or any of their component parts. A & R's Terms of Agreement and Financial Commitment for franchisees state that '[p]ayment of stock and all other expenses is managed by the Franchise Owner'.¹⁰⁹ A & R franchisees appear to have held the premises leases in their own names and to have dealt directly with the book suppliers. Thus the franchisor's failure would not result in franchisees losing the right to trade from their shops.

By 17 June 2011, the administrators had closed a further '42 [franchisor-owned] stores [leaving only] 19 company-owned stores alongside the 48-strong franchise network'.¹¹⁰ The franchisees were included in the discussions with two potential buyers of their franchise agreements. Options were presented to the A & R franchisees on 17 June. Most decided to join the Collins bookseller group¹¹¹ and the remainder joined an independent buying group called Leading Edge. Ongoing personal liability for their premises rental, independent of the success of their franchisor, would bear heavily on each franchisee's decision. The Applebys had every incentive to protect their very new investment.

Two features of the administration specific to A & R — gift cards and consumer warranties on merchandise purchased before the administration and normally returnable if faulty — serve to illustrate other considerations the franchisees needed to weigh up.

The February timing of the A & R administration meant that a high proportion of gift cards issued prior to Christmas and redeemable at any of the then 185 company-owned or franchisee-owned stores would not yet have been redeemed. Prior to the administration, Mrs Appleby pointed out that '[w]hen we sold a gift card, A & R took that money off us straight away. We didn't get that money back until someone used that gift card and then we had to claim it back'.¹¹² Once the franchisor was in administration, holders of

¹⁰⁷ Ferrier Hodgson, 'Second Meeting of REDgroup Creditors' (Press Release, 27 July 2011).

¹⁰⁸ Chris Zappone, 'Booksellers' Woes Worsen as Franchisees Seek to Defect', *The Sydney Morning Herald* (Sydney), 5 April 2011.

¹⁰⁹ Angus & Robertson, above n 56.

¹¹⁰ Smart Company, 'Meeting on Potential Sale of Angus & Robertson Underway' (17 June 2011) <<http://www.smartcompany.com.au/growth/economy/20478-20110617>>.

¹¹¹ Collins itself had been formed as a buyers' group after the Collins Booksellers franchisor failed in 2005 after selling books since 1929.

¹¹² 'Good Staff and Family Support See Stores Write Another Chapter', above n 59.

gift cards became unsecured creditors. The administrators advised as part of a series of updates about the progress of the administration that '[a]cceptance of gift cards is at the franchisee's discretion but the Administrators are under no obligation to repay you for any gift cards redeemed'.¹¹³ This operational change placed franchisees in an unenviable position. Whilst they became unsecured creditors for the face value of any unredeemed gift cards they also wanted to retain their customers. The Applebys 'honoured A & R gift cards for as long as they could, despite the fact that they were not legally bound to and despite making a loss on them'.¹¹⁴ The dilemma is illustrated in Figure 4.

The administrators' advice to franchisees not to replace faulty items or provide refunds to customers¹¹⁵ raised issues of warranties for faulty goods. Although this refusal would constitute a breach of the statutory warranties that provide consumer protection under the *CCA*, sch 2, *Australian Consumer Law*, ('*ACL*') ch 9, no claimant could

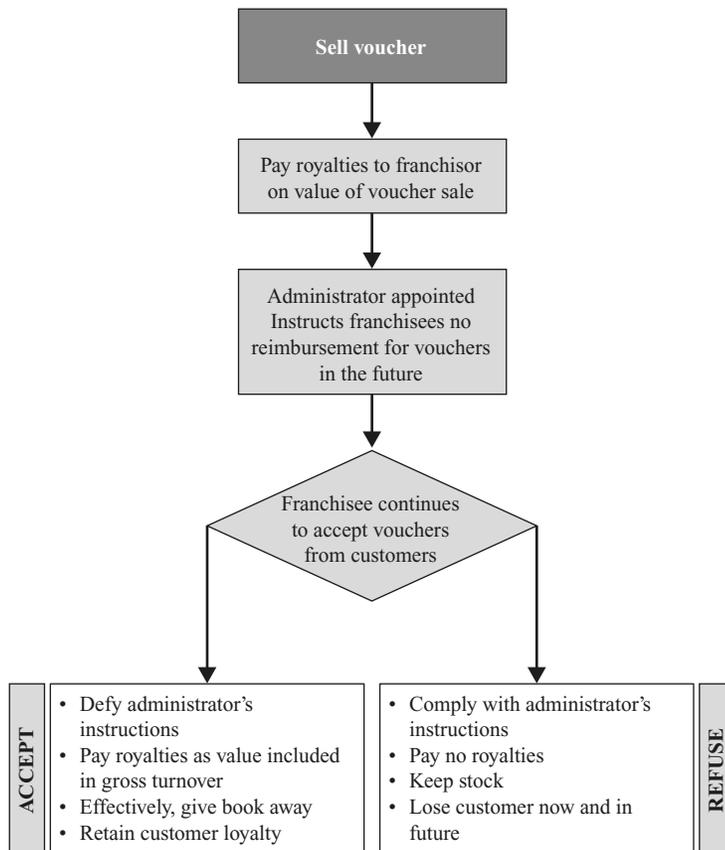


Figure 4: Gift card dilemma

¹¹³ Ferrier Hodgson, 'Frequently Asked Questions: Franchisees' (Circular, 17 February 2011) 2.

¹¹⁴ 'Good Staff and Family Support See Stores Write Another Chapter', above n 59.

¹¹⁵ Ferrier Hodgson, 'Frequently Asked Questions: Franchisees', above n 113, 2.

take action whilst the moratorium against creditors initiating or pursuing litigation was effective. In this regard the administrators' powers under the *CA* effectively 'trump' the consumer's rights under the *ACL*. Problems for franchisees go further than deciding how to manage customer loyalty. The provisions of the *CA* — including time limits and order of priority of payments — must be adhered to strictly by administrators. This includes a requirement that a second creditors' meeting must be held within 21 days of the appointment of the administrator. The court has discretion to consent to this meeting being held later if the administrators provide compelling reasons. In the REDgroup case (and similarly with Kleenmaid and Kleins), the administrators who had been appointed on 17 February 2011 were granted additional time to hold the second meeting of creditors. On 14 March 2011, Stone J ordered:

[p]ursuant to s 439A(6) of the *Corporations Act 2001* (Cth) (Act), ... the period within which the Administrators of the second plaintiffs must convene meetings of creditors of REDGroup Retail Pty Ltd and each other company names in the Schedule under s 439A of the Act [is] extended up to and including 18 September 2011.¹¹⁶

This enabled them to identify and negotiate with parties possibly interested in purchasing parts of the troubled business. The extended time frame placed the franchisees in limbo for 213 days from appointment of the administrator to the second creditors' meeting, 192 days (nearly 28 weeks) longer than the usual statutory period. This extended timeframe underscores the complexity of a franchisor administration and emphasises the franchisees' vulnerability.

We have seen through the A & R experience how franchisees move from being key stakeholders to being an incidental player if their franchisor becomes insolvent. Their investments are largely unprotected by the contracts, laws and the *Code* that helped inform their decision to enter the deal. We now turn to examine what franchisees may be able to do to protect their investment from the consequences of franchisor failure.

VI FRANCHISEES' EX ANTE AND EX POST RESPONSES TO FRANCHISOR INSOLVENCY

In general terms:

the effects of franchisor insolvency on the franchise ecosystem translates, upon insolvency, into myriad interests and competing claims among which the franchisee is the least protected. The interests of the franchisees are not protected and franchisees have no control over the business when the franchisor fails. Franchisees are subject to the decisions of the external controller.¹¹⁷

¹¹⁶ Re REDGroup Retail Pty Ltd ACN 108 801 127 and the Companies Listed in the Attached Schedule (All Administrators Appointed), No NSD 279/2011, Sealed Order, Federal Court of Australia, Sydney (14 March 2011).

¹¹⁷ Jenvey, above n 95, 2.

How, then, might franchisees insulate themselves *ex ante* and respond *ex post* to the announcement of its franchisors insolvency?

A Ex ante

Under the law of contract, prospective franchisees can ‘attempt to structure his or her affairs to ensure minimum personal liability and [maximum] flexibility in keeping or restructuring the business in the event the franchise business fails or alternatively the franchisor becomes insolvent’.¹¹⁸ Their ability to do so depends on the franchisor’s willingness to negotiate, the franchisor’s policies regarding matters such as whose name the premises lease is in, and how keen the franchisor is to make the sale. Franchisees would then be well advised to attempt to negotiate *ipso facto* clauses into all agreements they sign that depend on them continuing as a franchisee. Such clauses would be designed to provide franchisees with the right, but not the obligation, to terminate the agreement on the appointment of the franchisor’s administrator. It has also been suggested that a liquidated damages clause could be incorporated into contracts to avoid the franchisee being exposed to protracted litigation if they decide to terminate following the franchisor’s failure.

In the same way as indicators of a franchisee’s pending insolvency are easier for a franchisor to identify than the converse, franchisee failure is also easier for franchisors to pre-empt and navigate. A franchisor noticing its franchisee trading precariously may terminate the grant in reliance of a breach before the administrator is appointed. The franchisor does this in order to avoid dealing with the administrator. The franchisor will then be able to re-sell the former franchisee’s business and may avoid the risk of having it clawed back into the insolvency as a voidable preference. US franchise lawyer Craig Tractenberg recommends that ‘[i]f the franchise agreement is terminated before bankruptcy is filed, it is not protected by the automatic stay and the franchise is not property of the estate’.¹¹⁹ Failing this, the franchisor could negotiate to buy back the business from the franchisee’s administrator if the franchisor wants the site.

When conducting pre-purchase due diligence, a franchisee should take heed of the wording on the disclosure document ‘[t]his disclosure document contains *some of* the information you need in order to make an informed decision about whether to enter into a franchise agreement’¹²⁰ and think beyond it. Due diligence might include pursuing a range of lines of inquiry including: the direction of cash flow between franchisor and franchisees, market, risk of franchisor insolvency, and franchisor exit strategies.

It should not be assumed that the franchise is structured so the cash (eg royalties) flows from franchisee to franchisor. In some networks the franchisor receives the

¹¹⁸ Steven H Goldman, *Tackling Troublesome Insolvency Issues for Franchisees* (Unpublished, 2003) 3–6. Goldman’s paper outlines 10 specific strategies franchisees may attempt to put in place.

¹¹⁹ Tractenberg, above n 24, 7, citing *Moody v Amoco Oil Company*, 734 F 2d 1200 (7th Cir, 1984).

¹²⁰ *Code Annexure 1*, item 1.1 (e).

proceeds of each sale the franchisee makes direct from the franchisee's customer. The franchisor then remunerates its franchisee by paying commissions. In this situation the franchisee may be dependent on the franchisor for all of its cash flow. The intending franchisee should calculate how long its business could function if the franchisor did not pay in a timely way. With this risk clearly identified the franchisees should protect its own investment by seeking personal guarantees from the franchisor, and a right of setoff.

As mentioned previously, the *Code* prescribes mediation. Franchisees could seek security for moneys they become owed as a result of mediation. This would elevate them above unsecured creditors in the insolvency. It would also be prudent for franchisees entering mediation settlement agreements to require personal guarantees from the franchisor's directors if the agreement involved the franchisor paying money to the franchisees. This is because of the stay of proceedings¹²¹ the appointment of the administrator triggers vis-à-vis the franchisor.

Useful pre-commitment due diligence can be performed to determine whether the sector is viable in the medium term, or whether it is already saturated. US practitioner Cheryl Mullin suggests prospective franchisees

[g]oogle the franchisor with the key words 'earnings' and/or 'financial' and 'competitors' [and recommends that] 'earnings' and 'financial' often turns up statements that a principal or representation has made to the press. 'Competitors' will turn up articles addressing the competition. Often a franchise concept that is new to me is just one of many following a new craze or fad.¹²² If the barrier to entry is low, the concern is that even if the franchisee has a protected territory, it will soon be occupied by the competition. If there are competitors ... look at their [Franchise Disclosure Documents]¹²³ to compare fees, initial investment, and financial performance representations.¹²⁴

In addition to the disclosure requirements in the *Code*, franchisees are expected to seek advice from lawyers, accountants and business advisers. In correspondence to the Committee, the Franchise Council of Australia identified the main causes of franchisee failure to undertake effective due diligence as being:

- (1) aversion to incurring the costs of professional advice;
- (2) lack of understanding as to the value of professional advice or reluctance to seek advice;

¹²¹ *CA* ss 440D, 471(2).

¹²² For example the successful Boost Juice franchise in Australia was unsuccessfully mimicked by Pulp Juice Bars (owned by failed Signature Brands), and Nrgize (owned by failed Nrgize Australia Pty Ltd).

¹²³ Known as FDDs in the United States, this avenue is not available in Australia as there is no requirement that a franchisor's disclosure document be placed on a public data repository.

¹²⁴ Cheryl L Mullin (Mullin Law, Texas) Post on American Bar Association, *Forum on Franchising* (Wednesday 21 November 2012, 10.37 am) (access restricted).

- (3) a belief that they do not need assistance; and/or
- (4) ignorance of the *Code* requirement.¹²⁵

Due diligence has limitations. Part of the disclosure document may not be accurate. Some franchisors sign a statement of solvency when their business is insolvent. For example, franchisor Beach House Group ('BHG') was still accepting franchise fees from new franchisees in the third quarter of 2006. BHG's liquidator concluded that 'in my opinion the company became insolvent in 2005 and [knowingly remained] insolvent from that time'.¹²⁶ He also reported having

discovered an email prepared by one of the company directors dated 7 May 2007 admitting the company was insolvent and that the company should be wound up. There is no evidence to show the relevant director took any steps to prevent the company from incurring further debts.¹²⁷

Retail jewellery franchisor Kleins sold franchises while it was 'in [financial] trouble'.¹²⁸ Similarly, white goods franchisor Kleenmaid was described as being 'hopelessly insolvent'.¹²⁹ The liquidators were appointed on 15 May 2009 and reported 'each of the companies traded whilst insolvent prior to our appointment as Administrators. We are of the opinion that each of the companies in the group has been insolvent since March 2008'.¹³⁰ No franchisee could have uncovered damning financial information buried within the group accounts of franchisor-related insolvent Kleenmaid entities in time to avoid investing.

Franchisees must remember that franchisors also need exit strategies. Many are now at an age when they want to retire¹³¹ so the franchisor's succession planning is relevant. If possible the franchisees should make themselves aware of the franchisor's succession plan and should devote some thought to how the identified strategy might impact on the franchisees' business.¹³²

¹²⁵ Economic and Finance Committee, above n 1, 25.

¹²⁶ Cor Cordis, 'The Beach House Group Liquidators' Report to Creditors Pursuant to the *Corporations Act 2001* (Cth) s 508' (Liquidators' Report, 19 October 2009) 13.

¹²⁷ Ibid.

¹²⁸ Thomson, above n 15; see also Birkett Long Solicitors, above n 15.

¹²⁹ Richard Hughes (Liquidator, Deloitte), Comment by Member of Panel, Griffith University Franchise Forum, 2009.

¹³⁰ Deloitte, 'The Kleenmaid Group Liquidators' Report to Creditors Pursuant to the *Corporations Act 2001* (Cth) s 508' (Liquidators' Report, 25 August 2010).

¹³¹ In L Frazer, S Weaven and K Bodey, *Franchising Australia 2010* (Griffith University, 2010) 112, only 50 per cent of the franchisors surveyed were the founder of the original business.

¹³² For a preliminary discussion of this issue see Buchan, *Franchisees as Consumers*, above n 2, ch 7.

A final consideration is the power given to the ACCC under section 219 of the *CCA* to require advertisers (including franchisors seeking franchisees) to substantiate their claims.¹³³ If franchisees perceived that the general or franchisor-specific insolvency indicators mentioned earlier were present they could request that the ACCC use this power to require their franchisor to substantiate the claim that it is solvent. By exposing a franchisor's financial vulnerability through steps such as issuing of substantiation notice followed by an infringement notice or a public warning notice the ACCC could prompt a franchisor's proactive restructuring rather than allowing it to continue to fund itself through selling franchises, without which it would be insolvent.

B *Ex post*

The franchisees' ability to actively respond, *ex post* learning of the franchisor's insolvency, is limited. As noted earlier, the franchisor's insolvency does not constitute a breach of the franchise agreement. Under contract law the franchisee may argue that the franchisor's administration is an anticipatory breach that justifies the franchisee's termination. This places the franchisee at risk of a counterclaim by the administrator, but it is a strategy that some franchisees working under the commission agency model have used successfully.

Is it misleading and deceptive in breach of sch 2, s 18 of the *CCA* to continue to sell franchises when a franchisor suspects or knows it is insolvent, or is considering strategic insolvency? This possible action has never been litigated in the context of franchising, but in an analogous case concerning an employee *Moss v Lowe Hunt & Partners Pty Ltd*,¹³⁴ Katzmann J held that it was misleading or deceptive (under s 52 of the *Trade Practices Act 1974* (Cth), now sch 2, s 18 of the *CCA*) to describe a business as 'successful' when, without the continued support of its parent company (which itself had difficulties), it would be insolvent.

Bearing in mind the stay of proceedings mentioned previously, franchisees may turn their attention to solvent parties that facilitated their investment in the flawed franchisor. Some franchisors secure priority access to banks for prospective franchisees on the strength of the bank approving the franchise system. If the bank were to advance money to prospective franchisees on the strength of the franchise's credit approval ranking, is it misleading and deceptive of the financier if the franchisor was, in fact, the subject of a credit watch at the same bank at the time? It is arguable that this level of complicity could found an action for misleading and deceptive conduct under s 18 and/or unconscionable conduct under sch 2, s 20 of the *CCA* against the lender.

¹³³ See Stephen G Corones, *The Australian Consumer Law* (Lawbook, 2011) 495–7.

¹³⁴ [2010] FCA 1181.

VII POTENTIAL SOLUTIONS

The *CA* identifies the stakeholders whose interests merit special consideration. Franchise agreements and leases may be assets or liabilities to be dealt with accordingly, but franchisees, to the extent that they are not creditors, are excluded. Administrators thus have discretion as to how they treat franchisees. In the US some franchisor's administrators convene committees of franchisees. This creates a two-way information conduit and enables the administrators to gauge whether, for example, a group of franchisees is interested in buying the franchisor's business. It is a practice that should be adopted universally in Australian franchisor failures.

In relation to the franchise fee and money paid by franchisees to secure options to take up future franchises, the franchisor could be required to hold these on trust for the franchisee rather than shifting these amounts directly into a general revenue account. The franchisor would have the right to access amounts pro rata as the term of the grant passed milestones. The weight currently attributed to the value of franchisees' initial investment is not an appropriate reflection of their role within the franchise network, or of the risk they take. A more equitable solution than 'putting them in for \$1' would be to attribute to each franchisee a creditor weighting based on the written down value of their business at the time of the appointment, the residual value of their initial franchise fee (if, say the term was for 5 years and the franchisor failed after 1, then four fifths of the initial fee would be notionally owing by the franchisor as the franchisor has arguably been unjustly enriched by this amount), and any other amounts currently recognised as owing. Franchisees as a voting block at creditors' meetings could then, with the combined value of their investments, have an effective voice in the process.

The current order of insolvency priorities would thus need to be amended to accommodate the interests of franchisees. Because this would reduce the value of other creditors' claims, further research would be needed to ensure the *CA* priorities were equitably recalibrated.

VIII CONCLUSION

Franchisees invest debt-funded and equity capital in the franchisor. Traditional suppliers of debt finance such as banks have the opportunity to take security, and to price their loans accordingly. Traditional suppliers of equity finance, shareholders, have taken the risk of the entity becoming insolvent knowingly and on the basis of information supplied in a prospectus that has met rigorous statutory standards. Creditors, shareholders and employees enjoy clear rights in insolvency under the *CA*. Neither the law of contract nor the consumer protection regime offer effective relief to franchisees.

Any franchising industry code drafted under the *Small Business Commissioner Act 2011* (SA) should address today's highly sophisticated franchise networks. The simple model of a franchisor plus franchisees is rapidly becoming a thing of the past. The SA Review concluded that '[r]egulation ... cannot remove the possibility

of failure or guarantee success. Ultimately, a prospective franchisee's best protection against failure is educated, informed and conscientious due diligence'.¹³⁵ This must be read in the context of the amount of information that is inaccessible to franchisees (or accessible only at great cost) and of the accuracy of the amount disclosed by the franchisor. It must also be remembered that a franchisor may change the way it does business. It may decide to become insolvent after the franchisee has entered the system.

The franchise model contains fixable flaws, but legal solutions that focus on the time before the contract was formed cannot solve the conundrum of the franchisees' correct legal positioning in their franchisor's insolvency. It is also unrealistic to expect small business commissioners and consumer protection regulators to address the consequences of franchisor failure in isolation; teamwork is required. We must 'look beyond the boundaries of the legislative silos [to resolve] problems that cross boundaries'.¹³⁶ The team must include state and territory regulators of property rights, as well as insolvency policymakers and the corporations' regulator. The honey pot remains a source of sweet riches but the bear trap is becoming increasingly exposed. Its teeth are sharp.

¹³⁵ Economic and Finance Committee, above n 1, 25.

¹³⁶ Jenny Buchan *Franchisees as Consumers*, above n 2, 182 citing United Nations Commission on International Trade Law, *Legislative Guide on Insolvency Law* (2005) 19.

A RATIONAL APPROACH TO SENTENCING WHITE-COLLAR OFFENDERS IN AUSTRALIA

ABSTRACT

There are no overarching (and few settled) principles governing the sentencing of white-collar offenders. This is especially the situation in relation to the relevance of public opprobrium to the sentencing calculus and the manner in which employment deprivations stemming from the penalty impact on the sentence. To the extent that there is general convergence in the approach to sentencing white-collar offenders, the approach is often not sound. This is the case in relation to the minor sentencing discount accorded for previous good character, and the prevailing orthodoxy which assumes that offences targeted at major institutions, such as banks, meaningfully impair community confidence in such institutions. Fundamental reform of the manner in which white-collar offenders are sentenced is necessary in order to make this area of law more coherent and doctrinally sound. These reforms include providing a significant and pre-determined discount for restitution, reducing the weight given to general deterrence in the sentencing calculus, and providing a greater discount for previous good character and employment deprivations suffered as a direct result of the sentence. Further, crimes against individuals should be regarded as being more serious than those committed against large corporations or the public revenue. The article focuses on the existing law in Australia, however, the reform proposals and doctrinal analysis could be applied to all jurisdictions.

I INTRODUCTION

White-collar crime stands apart from other criminal offences. The differences are often exacerbated when it comes to sentencing white-collar offenders. Considerations that often distinguish white-collar offences from other types of crime include:

- white-collar offenders are not normally from socially-deprived backgrounds;
- white-collar offenders often do not have prior convictions;
- white-collar offences often involve a breach of trust or violation of some other moral virtue, such as loyalty;

* Professor, Dean and Head of School, School of Law, Deakin University.

** Lecturer, School of Law, Deakin University.

- the offences are normally well planned or may continue over a long period of time;
- it is often possible to fully remedy the resulting tangible harm through monetary restitution;
- there is no limit to the maximum benefit derived from the crime;
- the harm caused by the offence often goes beyond that inflicted on individuals and extends to financial institutions and markets;
- there are often non-criminal sequels to the conduct in question; and
- there is often a range of incidental sanctions which are suffered by white-collar offenders, including loss of reputation and reduction of future career prospects.

The unique features of white-collar offending have resulted in a fluid and unsettled jurisprudence in the sentencing of such offenders. This is unremarkable given the absence of legislative fiat dealing expressly with this criminological subset. This article examines the nature of white-collar crime with a view to providing a framework for a rational and consistent approach to sentencing white-collar offenders.

In part II of the article, we define ‘white-collar crime’. Part III analyses the current sentencing approach to white-collar criminals. In part IV, we suggest how the sentencing of white-collar criminals should be reformed.

The reforms we ultimately propose include: making restitution to the victim a stronger mitigating factor; providing a greater sentencing discount for previous good character; recognising the impact of incidental harms, such as diminished career prospects, in the sentencing process; and making the measure of harm caused by the offence the key determinant in sentence severity, consistent with the concept of proportionality.

II THE DEFINITION OF WHITE-COLLAR CRIME

There is no universally-accepted definition of white-collar crime despite the concept first being introduced over 80 years ago by Edwin Sutherland as ‘a crime committed by a person of respectability and high social status in the course of his occupation’.¹ This definition is forensically inadequate because notions such as ‘respectability’ and ‘social status’ are too obscure to be meaningful,² and

¹ This was the description given by Sutherland in a speech he delivered to the American Sociological Society in 1939: see J Kelly Strader, *Understanding White-Collar Crime* (LexisNexis, 2002) 1.

² Richard Quinney, ‘The Study of White-Collar Crime: Toward a Re-orientation in Theory and Research’ in Gilbert Geis and Robert F Meier (eds), *White-Collar Crime — Offenses in Business, Politics and the Professions* (Free Press, 1977) 285.

white-collar offending clearly transcends occupational or workplace transgressions.³

In a relatively recent analysis of white-collar crime in his book, *Lying, Cheating and Stealing: A Moral Theory of White-Collar Crime*, Stuart Green declines to attempt an exhaustive definition of what is encompassed by the concept. He notes the failure of sociologists to ascribe a consistent meaning to 'white-collar crime', observing that the conduct often concerns behaviour at the margins of the criminal spectrum. The diversity of conduct falling within the rubric of 'white-collar' leads Green to steer away from identifying semantic parameters, instead observing that such offending shares certain characteristics with other similar types of offending, though no single characteristic is essential.⁴

According to Green, key characteristics include: a diffused type of harm (often harming financial environments as opposed to identifiable individuals); wrongdoing that violates accepted norms (such as employee fidelity); and, a diminished role for mens rea.⁵ However, he concedes that many offences stand apart from this model. Thus, some offences which have a high mens rea, such as bribery and obstruction of justice offences, still fall within the white-collar crime rubric. The wrongs that Green identifies which are committed by white-collar criminals include breaches of trust, loyalty, dishonesty, deception and lying.⁶

The nebulous nature of white-collar crime is further illustrated by Arie Freiberg, who notes:

Discussion of the problem of sentencing [a] 'white-collar criminal' is plagued by the initial problem of identifying the subject matter. There is no discrete group of offences which can readily be identified as 'white-collar crime' ... [O]ver recent years, the phrase has been extended ... to cover any occupational deviance, whether by persons of high status or not, and violation of professional ethics. It would thus extend to cover the case of an academic who demands sexual favours in return for good grades. To some it has come to refer to almost any form of illegal behaviour other than conventional street crimes.⁷

³ For a discussion of the evolution of the white-collar terminology, see Kam Wong, 'From White-Collar Crime to Organizational Crime: An Intellectual History' (2005) *Murdoch University Electronic Journal of Law* 14.

⁴ Stuart P Green, *Lying, Cheating and Stealing: A Moral Theory of White-Collar Crime* (Oxford University Press, 2006) 18.

⁵ *Ibid* 34.

⁶ *Ibid*.

⁷ Arie Freiberg, 'Sentencing White-Collar Criminals' (Paper presented at the Fraud Prevention and Control Conference, Australian Institute of Criminology, Surfers Paradise, Queensland, 24–25 August 2000) 2 <http://aic.gov.au/media_library/conferences/fraud/freiberg.pdf> (citations omitted).

The United States Department of Justice gave a partial definition of white-collar crime in the following terms:

Nonviolent crime for financial gain committed by means of deception by persons whose occupational status is entrepreneurial, professional or semi-professional and utilizing their special occupational skills and opportunities; also, nonviolent crime for financial gain utilizing deception and committed by anyone having special technical and professional knowledge of business and government, irrespective of the person's occupation.⁸

Indeed, a workable definition of white-collar crime has proved so elusive that some commentators have suggested defining the phrase by what it is *not*. Thus, J Kelly Strader posits that a white-collar crime is one that does not:

- (a) necessarily involve force against a person or property;
- (b) directly relate to the possession, sale, or distribution of narcotics;
- (c) directly relate to organized crime activities;
- (d) directly relate to such national policies as immigration, civil rights, and national security; or
- (e) directly involve 'vice crimes' or the common theft of property.⁹

While definitions of white-collar crime can vary widely, it does not mean that the search for a definition should be abandoned. A working definition is crucial to the coherent analysis of existing jurisprudence, as well as to the process of informed decision-making and to any proposals for reform.

A white-collar crime involves an act which involves the taking of money or property (such as shares) or avoiding a legal obligation (such as a tax liability) without legal justification by an individual who is in a position of substantial influence regarding the relevant transaction. Examples of influence are where a bank employee transfers bank money into his or her account or where an individual submits a fraudulent tax return or deposits a false cheque or subverts the normal operation of the market system. It is further illustrated by paradigm instances of white-collar offending, which include:

- theft of company assets by company directors and employees (such as bankers);
- theft of client money by lawyers and accountants;
- insider trading and other market manipulation by people employed in industries associated with the financial markets;

⁸ Bureau of Justice Statistics, *United States Department of Justice, Dictionary of Criminal Justice Data Terminology* (2nd ed, 1981) 215, quoted in Strader, above n 1, 2.

⁹ Strader, above n 1, 2.

- complex tax fraud;
- corruption; and
- money laundering.

III CURRENT APPROACH TO SENTENCING WHITE-COLLAR OFFENDERS

A *General Matters Relevant to Sentence*

Sentencing is a complex activity and it is not feasible in an article of this size (and with its focus) to explain the key principles and rules.¹⁰ However, by way of background, we provide an overview of the structure of the sentencing law and the manner in which sentencing determinations are made as a backdrop to the remaining discussion.

Sentencing law and practice is not uniform throughout Australia. Each of the nine jurisdictions has distinctive statutes which guide sentencing decisions.¹¹ Sentencing law throughout Australia is, however, remarkably similar in the context of white-collar offending. This is because all of the sentencing statutes set out similar objectives of sentencing, in the form of general and specific deterrence, community protection, denunciation and rehabilitation.¹² In relation to each statutory scheme, there is no attempt to prioritise these sometimes conflicting objectives.¹³ The High Court of Australia has established that the key consideration in setting the penalty is the proportionality principle, which stipulates that the harshness of the sanction should match the seriousness of the offence.¹⁴

Moreover, the aggravating and mitigating considerations that inform the sentencing determination are largely universal. These factors are, principally, a manifestation of the common law, although some statutory schemes set out such matters in detail.¹⁵ Important factors that can aggravate or mitigate penalty include the level of harm caused by the offence, the offender's prior criminal record, remorse, the attitude of the victim (including victim impact statements), the level of planning

¹⁰ For an extensive analysis, see Richard G Fox and Arie Freiberg, *Sentencing: State and Federal Law in Victoria* (Oxford University Press, 2nd ed, 1999); Mirko Bagaric and Richard Edney, Thomson Reuters, *Australian Sentencing*, vol 1 (at Release 10); Geraldine Mackenzie and Nigel Stobbs, *Principles of Sentencing* (Federation Press, 2010).

¹¹ The main statutes that deal with sentencing in the respective Australian jurisdictions are as follows: *Crimes (Sentencing) Act 2005* (ACT); *Crimes Act 1914* (Cth) pt 1B ss 16–22A; *Crimes (Sentencing Procedure) Act 1999* (NSW); *Sentencing Act 1995* (NT); *Penalties and Sentences Act 1992* (Qld); *Criminal Law (Sentencing) Act 1988* (SA); *Sentencing Act 1997* (Tas); *Sentencing Act 1991* (Vic); *Sentencing Act 1995* (WA).

¹² See Bagaric and Edney, above n 10, [1-39101].

¹³ *Ibid.*

¹⁴ See further the discussion below.

¹⁵ See especially the *Crimes (Sentencing Procedure) Act 1999* (NSW) s 21A.

involved, the prevalence of the offence, the effect of the proposed sanction, hardship to others (especially the offender's family), any guilty plea, voluntary reparation, worthy social contributions, and assistance to the criminal justice system.¹⁶

All Australian sentencing schemes provide for the imposition of a similar range of sanctions. The least serious is a finding of guilt without any further harshness being imposed on the offender, and the most severe being a term of imprisonment.¹⁷

Further, the High Court has made clear that sentencing decisions are an ‘instinctive synthesis’ of all of the relevant variables as opposed to a mathematical calibration of each determinant that is relevant in a particular case.¹⁸ As a result, there is no single penalty which is objectively correct in any case. As was noted recently by the Victorian Court of Criminal Appeal in *Freeman v The Queen*:¹⁹

[i]t is a basic principle of sentencing law that there is no single correct sentence in a particular case. On the contrary, there is a ‘sentencing range’ within which views can reasonably differ as to the appropriate sentence.²⁰

Within this rubric, sentencing principles relating to white-collar crime are largely unsettled, but, in practice, there is nonetheless a number of widely-accepted and observed rules. Prior to examining the issue in some depth, it is possible to provide a framework by setting out some core matters.

White-collar offences are generally regarded as being committed principally for greed,²¹ thus, a paramount consideration in sentencing is the amount of money

¹⁶ Bagaric and Edney, above n 10, [1-42001], [1-60721].

¹⁷ Bagaric and Edney, above n 10, [1-501].

¹⁸ *Markarian v The Queen* (2005) 228 CLR 357; *Hili v The Queen* (2010) 242 CLR 520. The instinctive approach to sentencing means that discussions of sentencing which focus on increases and reductions in penalties are somewhat obscure, given that generally a mathematical or clear weighting is not ascribed to relevant variables. In this discussion we focus in particular on mitigating considerations. Although the extent to which any mitigating factors reduce the sanction is unclear, it is settled that where more than one mitigating factor is applicable, the factors operate in a cumulative manner to reduce sentence. There is no precise mathematical precision associated with this process other than the limitation that the factors cannot cumulate so significantly to result in a disproportionate sentence. This allows the courts to adopt a ‘rolled up’ or an ‘analytical approach’: see *R v Ehrlich* (2012) 219 A Crim R 415 (Adams J); *R v NP* [2003] NSWCCA 195 (17 July 2003); *R v El Hani* [2004] NSWCCA 162 (21 May 2004). The extent to which mitigating factors impact on sentence is made clearer by the fact that, as discussed below, certain considerations (namely, plea of guilty and assistance to authorities) attract a mathematical reduction to the penalty.

¹⁹ [2011] VSCA 214 (27 July 2011).

²⁰ *Ibid* [6].

²¹ See, eg, *R v Mercieca* [2004] VSCA 170 (10 September 2004).

involved.²² Other important considerations are the level of sophistication and planning of the offence²³ and whether or not a breach of trust occurred.²⁴ Offences committed over a long period of time are normally considered to be more serious.²⁵ That may be because long periods of time provide an offender the opportunity to desist from offending. Where an offender has desisted it is a strong mitigating factor, especially if coupled with an offender who voluntarily discloses the crimes to police.²⁶ The most important sentencing objective that the courts emphasise in sentencing white-collar criminals is general deterrence. It is mainly for this reason that white-collar crimes involving large amounts (roughly in the order of \$100 000) normally result in an immediate custodial term being imposed, irrespective of other mitigatory factors.²⁷

A good recent example of the general approach to sentencing white-collar offenders is set out in the comments of Warren CJ, Redlich J and Ross AJA in *DPP (Cth) v Gregory*²⁸ in the context of rejecting an appeal against sentence for a tax evasion offence. The passage is set out at some length to illustrate the complexity of the sentencing inquiry in such matters:

In seeking to ensure that proportionate sentences are imposed, the courts have consistently emphasised that general deterrence is a particularly significant sentencing consideration in white-collar crime and that good character cannot be given undue significance as a mitigating factor, and plays a lesser part in the sentencing process ... Moreover, general deterrence is likely to have a more profound effect in the case of white-collar criminals. White-collar criminals are likely to be rational, profit seeking individuals who can weigh the benefits of committing a crime against the costs of being caught and punished. Further, white-collar criminals are also more likely to be first time offenders who fear the prospect of incarceration.

In many if not most cases, imprisonment will be the only sentencing option for serious tax fraud in the absence of powerful mitigating circumstances. A sophisticated degree of planning accompanied by a lack of contrition should ordinarily lead to a more severe sentence of imprisonment. But despite the recognised importance of general deterrence, tax fraud has not always been as

²² Freiberg, above n 7, 9. See also, *Hoy v The Queen* [2012] VSCA 49 (7 March 2012).

²³ *R v Coukoulis* (2003) 7 VR 45; *R v Webber* (2000) 114 A Crim R 381; *R v Cameron* (1993) 171 LSJS 305 (SACCA).

²⁴ *R v Smith* (2000) 114 A Crim R 8.

²⁵ *R v Ralphs* [2004] VSCA 33 (9 March 2004) (fraud by a law clerk over a nine year period); *R v Grossi* (2008) 23 VR 500 (theft from employer lasting six years); *R v Galletta* [2007] VSCA 177 (30 August 2007) (theft from employer for nine years).

²⁶ *R v Lopez* [1999] NSWCCA 245 (12 March 1999); *R v Kittson* [2008] VSCA 77 (5 May 2008).

²⁷ *R v Whitnall* (1993) 42 FCR 512; *R v Nguyen* [1997] 1 VR 386; *DPP (Cth) v Carter* [1998] 1 VR 601; *R v Schwabegger* [1998] 4 VR 649, 654.

²⁸ *DPP (Cth) v Gregory* (2011) 250 FLR 169.

severely enforced as other forms of criminality. Over a decade ago this court, constituted by Winneke P, Brooking and Callaway JJA observed in *R v Nguyen and Phan* that the seriousness of the offence of defrauding the Commonwealth of income tax ‘has not always been sufficiently reflected in the sentence passed’ ...

A sentence imposed for fraud upon the taxation revenue, is intended to reaffirm basic community values that all citizens according to their means should fairly share the burden of the incidence of taxation so as to enable government to provide for the community, that the revenue must accordingly be protected, and that the offender should be censured through manifest denunciation. When these considerations are not reflected in the responses of the courts, the criminal justice system itself fails to achieve its objectives.²⁹

Several notable aspects emerge from the above passage. First, the preparedness of the court to look widely at the harm caused by a white-collar offence, in this case tax evasion, and focus not only on the obvious and immediate victim (the tax office) but also the ultimate victim (the general public). Second, the degree of planning and sophistication of the offence is important. Third, general deterrence is given considerable weight in the sentencing calculus because of the need to discourage similar offending. It is assumed that this will be especially effective in the case of those inclined to commit white-collar offences, because they do not normally have prior convictions and are likely to be rational agents who undertake a cost/benefit analysis prior to engaging in offending. Finally, it is noted that previously, in some instances, white-collar crimes were inadequately punished, which justifies an increase in current tariffs.

The cardinal role of general deterrence in relation to such crimes has been confirmed by numerous other authorities. For example, Sheller JA in *Director of Public Prosecutions (NSW) v Hamman*³⁰ stated:

General deterrence is a predominant consideration when sentencing for offences of defrauding the revenue. Appeal Courts have discussed and emphasised the seriousness of frauds committed to the detriment of the public revenue. Inevitably, the Australian system of tax collection depends upon the honesty of taxpayers and, in particular, upon their fully declaring in each year of income what their gross income is.³¹

The previous good character of the offender is recognised in the sentencing calculus. However, as was highlighted in *R v Gregory* above, the current sentencing orthodoxy maintains that this is the very feature of many white-collar offenders which enables them to commit such crimes, and accordingly, little weight is usually attributed to this factor.

²⁹ Ibid 182–4 [53], [54], [57] (citations omitted, emphasis added).

³⁰ (Unreported, New South Wales Court of Criminal Appeal, Sheller JA, 1 December 1998).

³¹ Ibid 6. See also *R v Bromley* (2010) 79 ATR 692, 697 [31]; *R v Nicholson; Ex parte DPP (Cth)* [2004] QCA 393 (22 October 2004) [20]; *R v Wheatley* (2007) 67 ATR 531.

In *R v Coukoulis*³² the role of good character in sentencing white-collar offenders was addressed by Ormiston JA in the context of a solicitor who committed a large-scale fraud (of more than \$8 million from over 40 people — mainly clients), where his Honour said:

[A]lthough the fact that the appellant has had *no prior convictions is of importance and must be recognised, these are the very circumstances in which he was able to deceive so many of his clients*. His very reputation, as a solicitor and generally in the community, enabled him to obtain the moneys he stole, to persuade his clients and others that he was always acting in their interests, to reassure them and to allow him to dissuade them from enquiring further as to the precise manner of their moneys' application. Implicit faith was, wrongly, placed in him in circumstances where, were the moneys obtained by an unqualified person or through a person with a lesser reputation, those depositors might well have been more cautious about protecting themselves against possible misuse of their moneys.³³

It is notable that, as in *Coukoulis*, above, assessing the harm done required the court to look again beyond the immediate victim, and to recognise the damage caused to the profession and the wider community in which the offender practised the profession. In evaluating the harm, the court also stated that the victims were caused greater distress because the offence involved a breach of trust. It was partly for this reason that the court stated that a heavy penalty was necessary to appropriately denounce the conduct.

In *R v Swift*³⁴ Nettle JA (with whom Vincent JA and Habersberger AJA agreed) noted that there is no tariff for white-collar offences, but that a breach of trust is a strongly aggravating factor, as is the financial vulnerability of the victim. Moreover, the Court gave little weight to factors personal to the accused (including loss of reputation and excellent prospects of rehabilitation) and rejected an appeal against a term of imprisonment of five years and three months for offending that spanned over two years, and totalled approximately \$1.5 million.

In *McMahon v The Queen*³⁵ the offender pleaded guilty to 38 counts of tax fraud and 42 counts of identity fraud, and was sentenced to a total effective term of six years imprisonment with a minimum term of four years. In rejecting his appeal against sentence, the New South Wales Court of Appeal again noted that while previous good character is relevant to the sentencing of white-collar offenders, it is pragmatically of little weight given the strong need for general deterrence.³⁶

The Court also rejected a submission that white-collar offenders should be met with a shorter non-parole period, primarily because of the growing seriousness

³² (2003) 7 VR 45.

³³ Ibid 59–60 [42] (emphasis added). See also *R v Kelvin* (2000) 46 ATR 1.

³⁴ (2007) 15 VR 497.

³⁵ [2011] NSWCCA 147 (22 June 2011).

³⁶ Ibid [76].

and visibility of such offences (in the community) and the fact that such offences are difficult and expensive to detect.³⁷ Even where a sentencing court is moved by an ‘exceptionally good prior character’ to wholly suspend a sentence of imprisonment, such as in *R v Pollard*³⁸ which involved electronic theft from an employer of about \$92 000 over three months, the Court of Criminal Appeal declared that such a sentence would be manifestly inadequate. It opined that ‘[t]he limited relevance of prior good character in formulating the length of an appropriate sentence cannot be transformed into a more compelling factor for the purposes of determining whether a sentence of imprisonment should be suspended’.³⁹

In *Stevens v The Queen*⁴⁰ the Court again noted the increasing seriousness with which such offending is viewed — even when a breach of trust is not involved — and the institutional damage that it can cause. That case involved a number of systematic deceptions in a bank, over a 16 year period which totalled approximately \$400 000. It noted that there can be no tariff for such offending, given the wide range of circumstances in which it can be committed, and the impact that a crime has on public confidence in a system is a strong aggravating consideration:

If public confidence in the integrity of the [electronic banking] system is to be maintained the courts have an obligation to ensure that when dishonest breaches of its security are identified the offenders are appropriately punished. Both personal and general deterrence are of particular significance in relation to these types of offences.⁴¹

Although there is no tariff, the courts have on occasions catalogued a large number of other white-collar facts and penalties to provide some level of consistency to the sentences imposed. Recent examples of this are *Pollock v Western Australia*,⁴² *Brennan v Western Australia*⁴³ and *Scook v The Queen*.⁴⁴

Voluntary restitution by the offender is consistently regarded by the courts as an important consideration. However, where restitution does occur, it does not automatically result in a significant discount because of the perception that an offender is ‘buying his or her way out of prison’.⁴⁵ In *R v Phelan*,⁴⁶ Hunt CJ at CL stated that

³⁷ Ibid [82]–[84]; See also *Hili v The Queen* (2010) 242 CLR 520, 538–41 [59]–[67].

³⁸ [2006] NSWCCA 405 (15 December 2006).

³⁹ Ibid [19]. It should be noted that the Court declined to allow the Crown appeal on the basis of the double jeopardy principle.

⁴⁰ (2009) 262 ALR 91.

⁴¹ Ibid 104 [79] (McClellan CJ at CL).

⁴² [2011] WASCA 133 (15 June 2011) [42].

⁴³ [2010] WASCA 19 (15 February 2010).

⁴⁴ (2008) 185 A Crim R 164.

⁴⁵ See *Kovacevic v Mills* (2000) 76 SASR 404, 421 [81] (Doyle CJ, Mullighan, Bleby and Martin JJ).

⁴⁶ (1993) 66 A Crim R 446.

offenders should not be able to purchase a lesser sentence, but then qualified this by stating: '[w]here there has been a substantial degree of sacrifice involved in the repayment, that is a matter which may properly be taken into account by way of mitigation'.⁴⁷

Thus, to the extent that voluntary restitution currently impacts upon a sentence, it is only in circumstances where the restitution constitutes a demonstrable hardship to the offender that it will provide a significant mitigation of penalty. The utilitarian benefits of reparation to the victim appear to be a secondary consideration.

B *Impact of Incidental Burdens and Hardships*

A particularly complex sentencing issue is the extent to which courts factor into the sentence hardships or burdens which directly stem from the offending, but which are not strictly part of the court-imposed penalty. These are referred to as the 'incidental burdens or hardships' flowing from the offence. They take two broad forms: those which are imposed by the courts and those not imposed by courts but are a direct consequence of the sentence. We consider them in that order.

There are three forms of curial deprivations. The first are confiscation orders. The second are restitution orders. The third are disqualifications from being involved with companies.⁴⁸ The main forms of non-curial hardships are:

- shame, embarrassment and social ostracism; and
- reduced employment and career prospects.

We first look at the current legal position regarding incidental curial deprivations.

1 *Confiscation Proceedings*

Confiscation proceedings are being increasingly used against offenders, especially in relation to drug and property offences.⁴⁹ Broadly, these proceedings can result in reclamation of property or money derived from the offence, or go further and strip

⁴⁷ Ibid 448. This approach was endorsed by Spigelman CJ (Price J agreeing) in *Thewlis v The Queen* (2008) 186 A Crim R 279, 280 [4]. See also *R v Conway* (2001) 121 A Crim R 177; *R v Berlinsky* [2005] SASC 316; *Stratford v The Queen* [2007] NSWCCA 279 (18 September 2007) [22]–[24]; *Chandler v The Queen* [2008] NSWCCA 240 (20 October 2008).

⁴⁸ For an extensive analysis of incidental curial deprivations, see Fox and Freiberg, above n 10, ch 6.

⁴⁹ The relevant statutory provisions are: *Confiscation of Criminal Assets Act 2003* (ACT); *Proceeds of Crime Act 2002* (Cth); *Confiscation of Proceeds of Crime Act 1989* (NSW); *Criminal Assets Recovery Act 1990* (NSW); *Criminal Property Forfeiture Act 2002* (NT); *Criminal Proceeds Confiscation Act 2002* (Qld); *Criminal Assets Confiscation Act 2005* (SA); *Crime (Confiscation of Profits) Act 1993* (Tas); *Confiscation Act 1997* (Vic); *Criminal Property Confiscation Act 2000* (WA).

the offender of property or assets regardless of whether they can be directly linked to the criminal activity or not. In essence, the general approach taken by sentencing courts is that the first type of orders will not result in mitigation of penalty, whereas the second will. The principles are set out in *R v McLeod*⁵⁰ as follows:

The obligation to disgorge the proceeds of crime is not a penalty. Disgorgement is necessary to prevent unjust enrichment. Forfeiture of the proceeds of crime has, nevertheless, been treated as a mitigating factor in some cases. Thus, it has been said that pecuniary penalty orders which relate entirely to profits from the unlawful activity constitute an additional punishment.

Disgorgement of benefits apart, forfeiture is relevant to penalty. At common law, forfeiture of lawfully-acquired property has generally been regarded as a mitigating factor in sentencing, since it places the offender in a worse position than he/she was before the commission of the offence. That is, forfeiture has a punitive or deterrent effect.

The sentencing principle of proportionality requires that the nature and extent of any forfeiture of property be considered in fixing the sentence. That is not to say that such orders are always to be viewed as warranting mitigation of penalty. It is necessary to consider whether the forfeiture will have a disproportionate or exceptional effect on the offender and may have a substantial deterrent effect.⁵¹

In that case, the Court also held that, in setting the penalty, the likelihood of future hardship in terms of exaction of property may be taken into account. This approach is reflected in s 320 of the *Proceeds of Crime Act 2002* (Cth).⁵² In *McMahon v The Queen*⁵³ the Court stated that repayment of money (in this case to the Australian Taxation Office) through a pecuniary penalty order is not in itself mitigatory,⁵⁴ however, consent to such an order can be taken into account as an indication of genuine contrition and remorse.⁵⁵

⁵⁰ (2007) 16 VR 682.

⁵¹ Ibid 685–6 [16]–[18]. In *R v Ford* (2008) 100 SASR 94 a similar approach was taken by Gray J (with whom Doyle CJ agreed).

⁵² This states:

A court passing sentence on a person in respect of the person's conviction of an indictable offence:

- (a) may have regard to any cooperation by the person in resolving any action taken against the person under this Act; and
- (b) must not have regard to any forfeiture order that relates to the offence, to the extent that the order forfeits proceeds of the offence; and
- (c) must have regard to the forfeiture order to the extent that the order forfeits any other property; and
- (d) must not have regard to any pecuniary penalty order, or any literary proceeds order, that relates to the offence.

⁵³ [2011] NSWCCA 147 (22 June 2011).

⁵⁴ As a consequence of s 320(d) of the *Proceeds of Crime Act 2002* (Cth).

⁵⁵ *McMahon v The Queen* [2011] NSWCCA 147 (22 June 2011) [72].

The situation is different in New South Wales and Western Australia, where s 24B of the *Crimes (Sentencing Procedure) Amendment Act 2010* (NSW) and s 8(3) of the *Sentencing Act 1995* (WA), respectively, preclude confiscation orders of any nature from being taken into account to reduce penalty.⁵⁶

2 Restitution

Courts in all jurisdictions have power to order restitution to victims of property offences.⁵⁷ In Victoria, for example, where goods have been stolen and a person found guilty of an offence connected with the theft, the court may order that the person return the stolen goods or the proceeds of their sale to their true owner.⁵⁸ As noted above in *Kovacevic v Mills*,⁵⁹ restitution in the context of property offences is a mitigating consideration, but is not necessarily a weighty factor.

3 Disqualifications

Individuals convicted of certain offences are disqualified — normally automatically — from being involved in the management of corporations for some period of time. ‘Managing corporations’ is defined expansively, not only prohibiting offenders from acting as company directors, but also from participating in corporate decision making or significantly affecting a company’s financial standing, or from communicating instructions or wishes to directors who might customarily act in accordance with those instructions. It, thus, operates as an effective total ban on any managerial involvement with a company.⁶⁰ The key provisions are contained in s 206B of the *Corporations Act 2001* (Cth) as follows:

- (1) A person becomes disqualified from managing corporations if the person:
 - (a) is convicted on indictment of an offence that:
 - (i) concerns the making, or participation in making, of decisions that affect the whole or a substantial part of the business of the corporation; or
 - (ii) concerns an act that has the capacity to affect significantly the corporation's financial standing; or

⁵⁶ In *Stock v The Queen* (2011) 206 A Crim R 574 it was held that confiscation proceedings even prior to the legislative changes in NSW were not normally mitigatory. This was also the view in *Greco v R* [2010] NSWCCA 268 (25 November 2010).

⁵⁷ *Crimes Act 1914* (Cth) s 21B; *Crimes (Sentencing) Act 2005* (ACT) ss 19–20; *Criminal Procedure Act 1986* (NSW) s 43; *Sentencing Act 1995* (NT) s 88; *Penalties and Sentences Act 1992* (Qld) ss 35, 194; *Criminal Law (Sentencing) Act 1988* (SA) s 52; *Sentencing Act 1997* (Tas) s 65; *Sentencing Act 1991* (Vic) s 84; *Sentencing Act 1995* (WA) ss 109–122.

⁵⁸ *Sentencing Act 1991* (Vic) s 84.

⁵⁹ *Kovacevic v Mills* (2000) 76 SASR 404, 421 [81].

⁶⁰ *Corporations Act 2001* (Cth) s 206A.

- (b) is convicted of an offence that:
 - (i) is a contravention of this Act and is punishable by imprisonment for a period greater than 12 months; or
 - (ii) involves dishonesty and is punishable by imprisonment for at least 3 months; or
- (c) is convicted of an offence against the law of a foreign country that is punishable by imprisonment for a period greater than 12 months.

The disqualification follows automatically upon conviction and the sentencing court is given no discretion.⁶¹ While not all white-collar offenders will suffer hardship equally as a consequence, Martin and Webster point out that any disqualification or disability ‘contributes significantly to the social stigma of the finding of guilt or conviction’.⁶² For those offenders whose occupation involved the management of a company, the disqualification presents an obvious and significant hardship. Courts have recognised this in some cases⁶³ but not in others.⁶⁴ Freiberg observes: ‘[t]he courts have been ambivalent on this issue, sometimes decreasing a sentence to take into account the additional detriment, and sometimes refusing to do so. The cases present no clear pattern’.⁶⁵

It should be noted that white-collar offenders can also be subject to disqualification orders consequent upon civil penalty proceedings, even where there are no criminal proceedings afoot.⁶⁶ The discretion resides with the prosecuting authority,⁶⁷ and there is no bar to pursuing a white-collar offender criminally after the conclusion of a civil penalty proceeding.⁶⁸

⁶¹ However, the Court has power upon application by the disqualified person to grant leave to manage a corporation in the future: *Corporations Act 2001* (Cth) s 206G.

⁶² J P Martin and D Webster, *The Social Consequences of Conviction* (Heinemann Educational, 1971) as cited in Fox and Freiberg, above n 10, 526.

⁶³ For example, specific reference to disqualification can be found in *R v Chan* (2010) 79 ACSR 189, 194 [20] (Forrest J); *DPP v Tang (aka Widjaja)* (Unreported, County Court of Victoria, Patrick J, 8 December 2011) 68.

⁶⁴ No reference to the statutory disqualification was made at all in *R v Hartman* (2010) 81 ACSR 121 or *R v Richard* [2011] NSWSC 866 (12 August 2011).

⁶⁵ Freiberg, above n 7, 12.

⁶⁶ *Corporations Act 2001* (Cth) ss 1317DA, 206C.

⁶⁷ The matters that inform the discretion whether to proceed by civil or criminal proceedings are unclear and the lack of transparency involved in this decision was noted by Higgins J in *R v Whitnall* (1993) 42 FCR 512, 518. See further Michelle Welsh, ‘The Regulatory Dilemma: The Choice Between Overlapping Criminal Sanctions and Civil Penalties for Contraventions of the Directors’ Duty Provisions’ (2009) 27 *Company and Securities Law Journal* 370; Michael Gething, ‘Do We Really Need Criminal and Civil Penalties for Contraventions of Directors’ Duties?’ (1996) 24 *Australian Business Law Review* 375.

⁶⁸ *Corporations Act 2001* (Cth) s 1317P.

4 *Public Opprobrium and Social Ostracism*

A common non-curial hardship stemming from white-collar offending is shame and embarrassment. The law is not settled on the impact that this should have on sentence. It was considered by the several members of the High Court in *Ryan v The Queen*,⁶⁹ but a majority of the Court did not endorse a clear position. Kirby and Callinan JJ stated that public opprobrium was a factor which could be taken into account to reduce the sanction imposed by the court, whereas McHugh J took the opposite approach. Gummow J did not canvass the issue, while Hayne J ‘substantially’ agreed with McHugh J.⁷⁰ Callinan J stated:

Of course the abuse of an office to commit a crime is greatly to be deplored but the crime of a person occupying an office of some prominence will often attract much greater vilification, adverse publicity, public humiliation, and personal, social and family stress than a crime by a person not so circumstanced. When these consequences are attracted they should not be ignored by the sentencing court.⁷¹

Kirby J agreeing with Callinan J stated:

[s]tigma [stemming from conviction] commonly add[s] a significant element of shame and isolation to the prisoner and the prisoner's family. This may comprise a special burden that is incidental to the punishment imposed and connected with it. If properly based on evidence, it could, in a particular case, be just to take such considerations into account in fixing the judicial punishment required.⁷²

McHugh J rejected the relevance of public opprobrium because:

First, it would seem to place a burden on the sentencing judge which would be nearly impossible to discharge. The opprobrium attaching to offences varies greatly from one offender and one offence to another. How a judge could realistically take such a matter into account is not easy to see ...

Secondly, the worse the crime, the greater will be the public stigma and opprobrium. The prisoner who rapes a child will undoubtedly be subject to greater public opprobrium and stigma than the prisoner who rapes an adult person.⁷³

In *R v Bunning*⁷⁴ the Court regarded it as mitigating that the offender ‘lost his reputation, his career [as a police officer] ... and suffered public humiliation’.⁷⁵

⁶⁹ (2001) 206 CLR 267.

⁷⁰ *Ibid* 313–14 [157].

⁷¹ *Ibid* 318–19 [177].

⁷² *Ibid* 304 [123]. In *McDonald v The Queen* (1994) 48 FCR 555 Burchett and Higgins JJ, at 564–5 [23], gave this considerable mitigating weight.

⁷³ *Ryan v The Queen* (2001) 206 CLR 267, 284–5 [53], [55].

⁷⁴ [2007] VSCA 205 (27 September 2007).

⁷⁵ *Ibid* [47].

The balance of authority indicates that shame can be a mitigating factor but that it generally carries little weight. In *Kenny v The Queen*,⁷⁶ Howie and Johnson JJ stated that public shame could be given some weight if it was so significant as to damage the person physically or psychologically. In *Einfeld v The Queen*⁷⁷ Basten JA (Hulme and Latham JJ agreeing on this issue) endorsed the position in *Kenny*, above, and stated that in that case two considerations could affect the manner in which public opprobrium factors into the sentencing calculus. The first was the offender's status as a former judge, which made the offence worse and gave rise to an increased level of public humiliation. Second, the offender used his previous position to advance his unlawful purpose.⁷⁸

An extensive analysis of the authorities was undertaken in *R v Nuttall; Ex parte Attorney-General (Qld)*⁷⁹ by Muir JA (Fraser and Chesterman JJA agreeing). The Court 'assumed' public opprobrium was relevant in light of the fact that it was not submitted that the sentencing judge failed to take it into account, but noted that public humiliation was of little weight given that it was inevitable:

The attainment of high public office brings with it public exposure and media scrutiny as well as power, fame and prestige. Criminal abuse of the office, if detected, will inevitably attract media attention and result in shame and distress to the offender and his family.⁸⁰

In part IV(D) below, we discuss the weight that public opprobrium should have in the sentencing calculus.

5 *Employment Deprivations: Dismissal or Loss of Opportunity to Work*

There is no generally agreed approach to the relevance of employment deprivations to sentence. A number of different approaches have been taken. In both *Kovacevic v Mills*⁸¹ and *G v Police*⁸² the sentence was mitigated to avoid damage to the offender's career prospects. In a similar vein, there have been a number of instances where sentences have been discounted because of consequential damage to career or prospects.⁸³ On the other

⁷⁶ [2010] NSWCCA 6 (12 February 2010).

⁷⁷ (2010) 266 ALR 598.

⁷⁸ Ibid 621 [98]–[101].

⁷⁹ (2011) 209 A Crim R 538.

⁸⁰ Ibid 553 [65].

⁸¹ (2000) 76 SASR 404.

⁸² (1999) 74 SASR 165.

⁸³ See *Moorhead v Police* [1999] SASC 243 (7 June 1999); *R v Richards* (1980) 2 Cr App R (S) 119; *Ryan v The Queen* (2001) 206 CLR 267; *Simmonds v Bureau of Customs* [2001] SASC 306 (29 August 2001); *Hook v Ralphs* (1987) 45 SASR 529; *R v McDermott* (1990) 49 A Crim R 105; *McDonald v The Queen* (1994) 48 FCR 555; *R v Purdon* (Unreported, Court of Criminal Appeal for New South Wales, Hunt, McInerney and Donovan JJ, 27 March 1997).

hand, in *R v Boskovitz*⁸⁴ and *Brewer v Bayens*⁸⁵ a sentence was imposed regardless of the effects on career or prospects, while in *R v Liddy [No 2]*⁸⁶ and *Hook v Ralphs*⁸⁷ the sentence was designed or calculated to destroy career or prospects.⁸⁸

The strongest statement regarding the supposed irrelevance of reduced employment prospects to the sanction which is imposed is found in the comments of McPherson JA in *R v Qualischefski*.⁸⁹ His Honour stated:

The applicant ... claims that a conviction for possession of cannabis will have dire consequences for him if it continues to be recorded. It will, he says, lose him his job as a computer operator with the Health Department, along with his career, his social position and his lifestyle. Those consequences are undoubtedly severe; but, if for that reason, appeals like this are allowed and recording of convictions set aside, the impact of the administration of justice will in the course of time be no less serious. *It will mean that we are sanctioning the division of offenders into two classes. There will be those with good jobs and careers, enviable social positions and prosperous life-styles. Their convictions will not be recorded for fear of the damage it may do them. Then there will be those without jobs, or career prospects, or with standards of living that are already depressed. In their case, convictions will be recorded. Such an outcome seems to me to be quite wrong and thoroughly indefensible. It smacks of privilege, and can only lead to the evolution of a special class of persons in society who are exempt from the full operation of the criminal law, at least at its lower reaches ...*

Most recently, in *R v Nuttall; Ex parte Attorney-General (Qld)*⁹⁰ Muir JA (Fraser and Chesterman JJA agreeing) took the view that: ‘the respondent’s loss of employment and lack of job prospects on his release are relevant considerations’.⁹¹ However, it is clear that the courts have failed to adopt a systematic or principled approach to the impact of likely employment deprivations on sentence.

⁸⁴ [1999] NSWCCA 437 (20 December 1999).

⁸⁵ (2002) 26 WAR 510. The appellant psychologist was convicted of solicitation consequent upon a random police sting operation. A conviction was recorded despite (or regardless of) the likely effects on his career, PhD studies and occupational contributions to the community.

⁸⁶ (2002) 84 SASR 231.

⁸⁷ (1987) 45 SASR 529.

⁸⁸ A sentence meant to put an end to a career may also be discounted *because* it has had that intended effect. In *R v Whitnall* (1993) 42 FCR 512, the sentence was increased as a consequence of the defendant’s career.

⁸⁹ (Unreported, Queensland Court of Appeal, Fitzgerald P, McPherson JA, Ambrose J, 12 August 1994) 4 (emphasis added). This case is discussed in Andrew West, ‘Prospective Loss of Employment as a Factor in Mitigation of Penalty’ (1996) 16 *Queensland Lawyer* 157.

⁹⁰ [2011] 2 Qd R 328.

⁹¹ *Ibid* 343 [59].

C Summary of Sentencing Principles Relevant to White-Collar Offences

The above analysis demonstrates that sentencing principles and practices are not uniform or well-settled in relation to white-collar criminals. However, some key themes emerge from the decisions:

- The type of conduct which constitutes a white-collar crime varies widely, hence, it is not feasible for a sentencing tariff to be developed.
- A key consideration in determining offence severity for white-collar crimes is the amount of money or value of property involved.
- General deterrence is commonly identified as the paramount objective in sentencing white-collar offenders. This serves to increase the penalty — often considerably — at the expense of all other sentencing factors.
- Crimes that are well-planned and committed over a long period of time are often punished more heavily.
- Breach of trust is a strong aggravating factor.
- Restitution of the amount taken is a mitigating factor but, generally, does not carry considerable weight.
- Offences against individuals are sometimes regarded as being more serious, but there is no clear acceptance of this proposition. It is assumed that offences against institutions can damage institutional and investor confidence, threaten the revenue or endanger the community good generally.
- Previous good character carries little mitigating weight, given that this is said to be both typical and facilitative of white-collar offending.
- The fact that an offender has made a worthwhile contribution to the community is often of little mitigating weight.
- Good family background is also of little mitigating weight.
- Extra-curial harm suffered as a result of being convicted and punished, in the form of loss of reputation, social and public humiliation and embarrassment and reduction in employment prospects, normally has little weight.
- Penalties imposed incidentally or consequently upon conviction in the form of confiscation, restitution and/or disqualification orders may or may not be taken into account, depending on the jurisdiction and the reason for the penalty.

IV FLAWS WITH THE CURRENT SENTENCING APPROACH

There are several flaws with the current approach to sentencing white-collar offenders. Unsupportable assumptions underlying current sentencing practices and inconsistency in approach are evident. We start with a consideration of the proportionality principle and its implications for sentencing practice for white-collar offending.

*A The Principle of Proportionality:
Offences Against Individuals Are More Serious*

The key determinant in the sentencing of white-collar offenders, as with all offenders, is the principle of proportionality.⁹² In crude terms, this means that the punishment must fit the crime. This is underpinned by the broader principle that benefits and burdens should be distributed with regard to, and commensurate with, a person's merit or blame. Proportionality operates to restrain not only sentences that are too heavy, but also those that are too light.⁹³ A clear statement of the principle of proportionality is found in the High Court case of *Hoare v The Queen*:⁹⁴

a basic principle of sentencing law is that a sentence of imprisonment imposed by a court should never exceed that which can be justified as appropriate or proportionate to the gravity of the crime considered in light of its *objective* circumstances.⁹⁵

The proportionality principle has two limbs: the harm caused by the offence and the level of pain inflicted by the punishment. The requirement of proportionality is satisfied if these limbs are aligned.

In *Veen v The Queen*⁹⁶ and *Veen v The Queen [No 2]*⁹⁷ the High Court stated that proportionality is the primary aim of sentencing. Proportionality is considered so important that it cannot be trumped even by the goal of community protection, which

⁹² Preserving proportionality between the seriousness of the offence and the severity of the sentence is a principle adhered to by all western legal systems: see Richard S Frase, 'Comparative Perspectives on Sentencing Policy and Research' in Michael Tonry and Richard S Frase (eds), *Sentencing and Sanctions in Western Countries* (Oxford University Press, 2001) 259, 261, cited in Ruth Kannai, 'Preserving Proportionality in Sentencing: Constitutional or Criminal Issue' (Paper presented at 18th International Conference of the International Society for the Reform of Criminal Law, Montreal, August 2004) <<http://www.isrcl.org/Papers/2004/Kannai.pdf>>.

⁹³ *R v Hernando* (2002) 136 A Crim R 451, 459 (Heydon JA).

⁹⁴ (1989) 167 CLR 348.

⁹⁵ *Ibid* 354 (Mason CJ, Deane, Dawson, Toohey and McHugh JJ) (emphasis in original).

⁹⁶ (1979) 143 CLR 458, 467 (Stephen J), 478 (Jacobs J).

⁹⁷ (1988) 164 CLR 465, 472 (Mason CJ, Brennan, Dawson and Toohey JJ).

at various times has also been declared as the most important aim of sentencing.⁹⁸ Thus, for dangerous offenders, while community protection remains an important objective, at common law it cannot override the principle of proportionality. It is for this reason that preventive detention is not sanctioned by the common law.⁹⁹

Proportionality has also been given statutory recognition in most Australian jurisdictions. For example, in Victoria, the *Sentencing Act 1991* (Vic) provides that one of the purposes of sentencing is to impose just punishment,¹⁰⁰ and that in sentencing an offender the court must have regard to the gravity of the offence¹⁰¹ and the offender's culpability and degree of responsibility.¹⁰² The *Sentencing Act 1995* (WA) states that the sentence must be 'commensurate with the seriousness of the offence',¹⁰³ and the *Crimes (Sentencing) Act 2005* (ACT) provides that the sentences must be 'just and appropriate'.¹⁰⁴ In the Northern Territory and Queensland, the relevant sentencing statutes provide that the punishment imposed on the offender must be just in all the circumstances,¹⁰⁵ while in South Australia the emphasis is upon ensuring that 'the defendant is adequately punished for the offence'.¹⁰⁶ The need for a sentencing court to 'adequately punish' the offender is also fundamental to the sentencing of offenders for Commonwealth matters.¹⁰⁷ The same phrase is used in New South Wales.¹⁰⁸

The courts have not attempted to exhaustively define the factors that are relevant to proportionality. The broad approach taken to this problem is to adopt the principle that the upper limit for an offence depends on its objective circumstances. However, some factors have been positively identified as relevant to offence seriousness. These include the consequences of the offence (including the level of harm), the victim's vulnerability, the method of the offence, the offender's culpability (which turns on such factors as the offender's mental state),¹⁰⁹ and the level of sophistication involved.¹¹⁰

⁹⁸ For example, see *Chanon v The Queen* (1978) 33 FLR 433; *R v Valenti* (1980) 2 A Crim R 170, 174; *DPP (Cth) v El Karhani* (1990) 21 NSWLR 370, 377.

⁹⁹ *Chester v The Queen* (1988) 165 CLR 611, 618.

¹⁰⁰ *Sentencing Act 1991* (Vic) s 5(1)(a).

¹⁰¹ *Ibid* s 5(2)(c).

¹⁰² *Ibid* s 5(2)(d).

¹⁰³ *Sentencing Act 1995* (WA) s 6(1).

¹⁰⁴ *Crimes (Sentencing) Act 2005* (ACT) s 7 (1)(a). The *Sentencing Act 1997* (Tas), however, does not refer to the principle of proportionality.

¹⁰⁵ *Sentencing Act 1995* (NT) s 5 (1)(a); *Penalties and Sentences Act 1992* (Qld) s (9)(1)(a).

¹⁰⁶ *Criminal Law (Sentencing) Act 1988* (SA) s 10(1)(k).

¹⁰⁷ *Crimes Act 1914* (Cth) s 16A(2)(k).

¹⁰⁸ *Crimes (Sentencing Procedure) Act 1999* (NSW) s 3A(a).

¹⁰⁹ For example, whether it was intentional, reckless or negligent.

¹¹⁰ New South Wales Law Reform Commission, *Sentencing*, Discussion Paper No 33 (1996) 62–4.

In terms of property offences, the harm comes in two main forms. The first is financial loss to the victim. This is dependent on the value of property taken from the victim, and the means of the victim. To this end, high value offences will usually cause more suffering than small ones, and real individuals are usually more harmed than large institutions, which have a greater capacity to recover losses or build them into their financial planning. Thus, crimes committed against individuals, especially those who are financially vulnerable or fragile (ie the most poor, the unemployed or financially struggling), cause more direct and much greater harm than crimes committed against the revenue or large corporations. An individual's capacity to recover is often limited and their interests are demonstrably set back by such crimes.

Accordingly, the law should be reformed to reflect the fact that white-collar offences involving the taking of money from individuals should be punished more heavily. Not only would this more clearly adapt the proportionality principle, but it would also enhance the doctrinal consistency of sentencing principle. Victim vulnerability is an entrenched aspect of sentencing practice. It is an established aggravating factor in relation to a range of offences, including assaults against the elderly¹¹¹ young,¹¹² and people with an intellectual disability.¹¹³

The second type of harm that is caused by white-collar offending is damage to institutional integrity and investor confidence. Depending on the crime, harm is normally caused to institutions, by insider trading offences, or to the body politic itself, by revenue offences. Sentencing courts often use the collective 'community interests' as a catch-all phrase to describe the victim. It could be argued that where the victim is *everyone* as opposed to *someone*, this type of offending is as serious, or perhaps even worse, than crimes against individuals. Former Federal Court judge Raymond Finkelstein, who heard many of the largest white-collar criminal trials and pleas, publicly adheres to this view.¹¹⁴

However, damage to institutional integrity is speculative at best. There is no evidence of a correlation between, say, share market activity and insider trading convictions or bank deposits and bank fraud. If such a relationship did exist, presumably, the direct victims of such crimes would demonstrate the greatest reduction in confidence in the financial system. In the United Kingdom, a scandal involving a pension-fund fraud committed by former Member of Parliament Robert Maxwell, which affected

¹¹¹ *DPP (Vic) v Grabovac* [1998] 1 VR 664; *R v Eisenach* [2011] ACTCA 2 (17 February 2011); *Royer v Western Australia* (2009) 197 A Crim R 319; *R v El-Chammas* [2009] NSWCCA 154 (2 June 2009).

¹¹² *R v AJW* (2001) 80 SASR 246; *Ollis v The Queen* [2011] NSWCCA 155 (14 July 2011); *Rioli v The Queen* [2010] NTCCA 13 (18 October 2010); *R v Eisenach* [2011] ACTCA 2 (17 February 2011).

¹¹³ For instance, see *R v Grech* [1999] NSWCCA 268 (6 September 1999) [37] (Carruthers AJ).

¹¹⁴ Patrick Durkin, 'Top Judge Takes Aim at White-Collar Criminals', *The Australian Financial Review* (Sydney), 10 June 2011.

25 000 individuals, led to a small study on the attitudinal effects of the crime on 25 of those individuals.¹¹⁵ Spalek concluded that:

The study reported in this paper illustrates that in some cases of fraud, victims may not be ‘duped investors’, but rather may distrust particular agents prior to any crime occurring, and may therefore be engaging in risk avoidance strategies. As a result, becoming the victim of a financial crime may not necessarily lead to individuals avoiding the financial system in general, because an integral part of their trust may be acknowledging that as investors they run risks.

The absence of a correlation between financial crimes and trust in the political and economic arenas is supported by research conducted elsewhere.¹¹⁶ Consequently, reform is required such that criminal sanctions (which have real consequences) are not made more severe in order to reflect imaginary harm to ‘the community’ without tangible evidence in support of that approach.

Arguably, institutional integrity has a relevance beyond investor confidence in market systems. Individuals may well be able to operate in a corrupt system, however, in a competitive global market, less money may flow into countries which have corrupt markets.¹¹⁷ Even this broad consideration of the meaning of institutional integrity does not justify more severe punishment for white-collar criminals. There is no evidence that collective market honesty and transparency is a principal driver of the international flow of funds — as opposed to where investors feel they can maximize their return. Even if a strong link between market integrity and international money flows is established, the connection between the negative impact on the entire market and any single criminal act is likely to be so minor that it would violate proportionalism to meaningfully increase sanction severity for this reason.

B Restitution Should Be a Strong Mitigating Factor — It Should Reduce the Penalty By Up to 30 Per Cent

As noted above, one limb of the proportionality principle is the harm caused by the offence. This has considerable implications for the manner in which restitution by white-collar offenders should be treated. As we have seen, the courts place some weight on restitution, but normally it is not a cardinal sentencing consideration.

¹¹⁵ Basia Spalek, ‘White-Collar Crime Victims and the Issue of Trust’ (Paper presented at British Criminology Conference, Leicester, July 2000) <<http://www.britsocrim.org/v4.htm>>.

¹¹⁶ The general conclusion in these studies is that white-collar crimes have little, if any, effect on trust: see, for example, John G Peters and Susan Welch, ‘The Effects of Charges of Corruption on Voting Behavior in Congressional Elections’ (1980) 74 *American Political Science Review* 697; Michael Mills and Elizabeth Moore, ‘The Neglected Victims and Unexamined Costs of White-Collar Crime’ (1990) 36 *Crime and Delinquency* 408; Neal Shover, Greer Litton Fox and Michael Mills, ‘Long-Term Consequences of Victimization by White-Collar Crime’ (1994) 11 *Justice Quarterly* 75.

¹¹⁷ We thank the anonymous referee for this observation.

The main rationale is that if restitution were given more prominence in sentencing, it would theoretically enable wealthy offenders to ‘buy their way out of prison’. There is some force to that argument and, in principle, it is an undesirable outcome. However, the practical consequence of adhering to that view is damage to victims. It means that offenders are not provided with any pragmatic incentive to repay victims and thus redress the harm they have caused.

Undeniably, non-restitution of relatively large sums of money, particularly to the financially vulnerable, can have a devastating impact on lives: in terms of health, enjoyment and longevity.¹¹⁸ This is a high price for victims to pay for doctrinal soundness. Thus, the harm from a crime can either be perpetuated, or it can be controlled, by acts of restitution. There is no easy way to resolve whether we should opt for ‘pure principle’ in the form of not rewarding offenders for paying back sums stolen from victims (a deontological perspective), or pursue good outcomes for victims and encourage those acts through sentencing discounts (a utilitarian perspective). However, as a general rule, speculative benefits should not be preferred over concrete benefits. This is certainly the manner in which the sentencing system resolves other similar tensions. It is readily observable in other contexts, such as the guilty plea discount and the discount for giving evidence against co-offenders.

All accused are entitled to plead not guilty and make the prosecution prove its case. In principle, offenders who plead not guilty to a crime should on no account be punished more heavily than those who plead guilty (apart from the extent to which a guilty plea is indicative of remorse). Yet, for pragmatic reasons, sentencing law provides a large discount to offenders who plead guilty. This is for no higher or more virtuous reason than without the discount, the court system would become clogged.

The High Court of Australia in *Cameron v The Queen*¹¹⁹ approved of the plead guilty discount and, in the process, the majority of the Court rejected a number of arguments against the discount, including that it constitutes a form of discrimination against offenders who elect to pursue their ‘right’ to a trial.

In all Australian jurisdictions, accused who plead guilty receive a sentencing discount.¹²⁰ This is so, irrespective of whether the plea is coupled with remorse. For example, in *R v Morton*¹²¹ the Victorian Court of Criminal Appeal stated that:

¹¹⁸ Barbara O’Neill et al, ‘Negative Health Effects of Financial Stress’ (2005) 51 *Consumer Interests Annual* <<http://www.personalfinancefoundation.org/research/efd/Negative-Health-Effects-of-Financial-Stress.pdf>>.

¹¹⁹ *Cameron v The Queen* (2002) 209 CLR 339.

¹²⁰ *Crimes Act 1914* (Cth) s 16A(2)(g); *Crimes (Sentencing) Act 2005* (ACT) s 35(2); *Penalties and Sentences Act 1992* (Qld) s 13(4); *Sentencing Act 1995* (NT) s 5(2)(j); *Crimes (Sentencing Procedure) Act 1999* (NSW) s 22(1); *Criminal Law (Sentencing) Act 1988* (SA) s 10(g); *Sentencing Act 1991* (Vic) s 5(2)(e); *Sentencing Act 1995* (WA) s 8(2).

¹²¹ [1986] VR 863.

A plea of guilty may be taken into account regardless of whether or not it is also indicative of some other quality or attribute such as remorse ... A court may always take a plea of guilty into account in mitigation of sentence even though it is solely motivated by self-interest.¹²²

The main reason for the discount is purely the utilitarian benefit in the form of clearing court backlogs. Gaudron, Gummow and Callinan JJ in *Cameron v The Queen*¹²³ stated:

[Australian courts] have taken the pragmatic view that giving sentence ‘discounts’ to those who plead guilty at the earliest available opportunity encourages pleas of guilty, reduces the expense of the criminal justice system, reduces court delays, avoids inconvenience to witnesses and prevents the misuse of legal aid funds by the guilty.¹²⁴

The guilty plea discount is one of only two situations where a stated numerical discount is usually applied by the courts, despite the endorsement of the instinctive synthesis approach to sentencing.¹²⁵ The reason is to underline to the accused the reality of the discount and, hence, encourage greater pragmatism by them.

The normal range of the discount is between 10 per cent and about 30 per cent, depending on the circumstances of the case. In several jurisdictions it is either conventional or a statutory requirement to indicate the size of the discount.¹²⁶ In *R v Thomson*,¹²⁷ the New South Wales Court of Criminal Appeal issued a guideline

¹²² Ibid 867. This approach was also adopted by Spigelman CJ (with whom other members of the Court agreed) in *R v Thomson* (2000) 49 NSWLR 383, 411 [115] who stated that there are ‘benefits to the criminal justice system as a whole’ that result from a guilty plea. At 412, [122], his Honour further noted that the ‘public interest served by encouraging pleas of guilty for their utilitarian value is a distinct interest’.

¹²³ *Cameron v The Queen* (2002) 209 CLR 339.

¹²⁴ Ibid 73–4 (emphasis added) (citations omitted).

¹²⁵ This was most recently endorsed by the High Court in *Hili v The Queen* (2010) 242 CLR 520.

¹²⁶ In New South Wales and Queensland, the Court must indicate if it *does not* award a sentencing discount in recognition of a guilty plea: *Crimes (Sentencing Procedure) Act 1999* (NSW) s 22(2) and *Penalties and Sentences Act 1992* (Qld) s 13(3)). In South Australia, Western Australia and New South Wales, the courts often specify the size of the discount given. In Victoria, s 6AAA of the *Sentencing Act 1991* (Vic) states that when courts provide a discount for a plea of guilty, they must specify the sentence that would have been given in the absence of that discount. An appeal does not lie against that ‘notional sentence’. See *R v Burke* (2009) 21 VR 471. Also see *Giordano v The Queen* [2010] VSCA 101 (7 May 2010) [45]. There has been some judicial comment as to the artificiality of s 6AAA given the instinctive synthesis that produces the actual sentence. See *Scerri v The Queen* (2010) 206 A Crim R 1, 5 [23]–[25]; *R v Flaherty [No 2]* (2008) 19 VR 305. The rationale and size of the typical discount in Victoria is discussed in *Phillips v The Queen* [2012] VSCA 140 (29 June 2012).

¹²⁷ (2000) 49 NSWLR 383.

judgement stating that a guilty plea will generally be reflected in a 10 to 25 per cent discount on sentence, depending on how early the plea is entered and the complexity of the case.¹²⁸ This suggested range relates only to the utilitarian value of a guilty plea to the criminal justice system and does not include additional discounts that may be available — for example, where the guilty plea may be said to evidence remorse. This now has a legislative basis.

Section 17 of the *Criminal Case Conferencing Trial Act 2008* (NSW) states that an early plea attracts a discount of up to 25 per cent, and late pleas can obtain a discount of up to 12.5 per cent. In *Lee v The Queen*¹²⁹ it was held where the plea was taken on the second day set for trial that a 12.5 per cent discount was appropriate. The co-offender received a 20 per cent discount for pleading on arraignment and it was held that the difference was appropriate.¹³⁰ In Western Australia, the discount often ranges from 20 to 35 per cent under the state's 'fast track system'.¹³¹ The Western Australian Court of Appeal rejected submissions that a full 'discount of the order of 30 per cent will *automatically* be afforded for a fast-track plea of guilty without more'.¹³² There is no requirement to quantify the discount in Western Australia.¹³³ In South Australia the common range is between 15 to 25 per cent, with 25 per cent regularly given for an early plea of guilty.¹³⁴

Providing assistance to authorities is treated in a similar way to guilty pleas, particularly where it results in the detection and prosecution of other offenders. It is important to note that, as with the guilty plea discount, this benefit is given independent of any reasons or remorse that might be demonstrated by assisting authorities. Criminals, in principle, should not be dealt with less severely because they opportunistically decide to give evidence against co-offenders. However, as a matter of public policy, the law encourages those involved in criminal behaviour to betray the confidence reposed in each other by providing a significant discount at the sentencing stage of the criminal justice system.¹³⁵ This is especially apposite given that it often places the individual in personal danger.¹³⁶

¹²⁸ See also *Charkawi v The Queen* [2008] NSWCCA 159 (4 July 2008); *R v Bugeja* [2001] NSWCCA 196 (11 May 2001).

¹²⁹ [2011] NSWCCA 169 (28 July 2011).

¹³⁰ The same discount was accorded in *Nakhla v The Queen* [2011] NSWCCA 143 (24 June 2011).

¹³¹ See, eg, *Trescuri v The Queen* [1999] WASCA 172 (10 September 1999); *Deering v Western Australia* [2007] WASCA 212 (17 October 2007).

¹³² *Cameron v The Queen* [2002] WASCA 81 (22 March 2002) [19].

¹³³ *McLean v Western Australia* [2011] WASCA 60 (16 March 2011) [57].

¹³⁴ *Fisher v Police* (2004) 154 A Crim R 511.

¹³⁵ *Malvaso v The Queen* (1989) 168 CLR 227, 239 (Deane and McHugh JJ).

¹³⁶ *R v Barber* (1976) 14 SASR 388, 390 (Bray CJ). See also *DPP (Cth) v AB* (2006) 94 SASR 316.

Assistance to law enforcement officials enjoys recognition in a number of statutory regimes.¹³⁷ In terms of the size of the discount that is available, it has been held that the discount for a plea of guilty and assistance to authorities should be up to 50 per cent.¹³⁸

Thus, when the legal system wishes to strongly encourage a course of conduct, the law is willing, with arithmetical clarity, to provide a significant and clear sentencing discount in order to encourage that conduct. This occurs despite the courts' steadfast adherence to the instinctive synthesis theory, and despite the questionable morality of permitting resource allocation to dictate the punitive consequences of criminal conduct. For far better reasons, the same approach should be adopted in respect of restitution to victims of crime.

Property offences are one of the few crimes where it is possible to restore the victim, economically, to the same position as before the offence. Where restitution does not occur, victims can be devastated — especially when they are individuals. There is a measurable difference in terms of harm caused between white-collar crimes that are restituted and those which are not. This should be reflected in the sentencing calculus. White-collar offenders who provide full restitution should get a discount of up to 30% — proportionally less for partial restitution.

It is accepted that this will place property offenders in a different situation from other offenders, but does not mean that they are given an opportunity to buy their freedom. It simply reflects the different nature of their crime: the targeting of commodities as opposed to other human interests. Moreover, the lives of victims should not be sacrificed in the fanatical pursuit of the dubious principle that offenders should not buy their way out of gaol. Principle often yields to the practical imperatives of the criminal justice system;¹³⁹ the lives of individual victims are no less important.

This approach entails that offenders who wish to but cannot make restitution, because they have dissipated the proceeds of their offending (for example, because of drug or gambling addictions), will receive more severe penalties than offenders

¹³⁷ See *Penalties and Sentences Act 1992* (Qld) s 9(2)(i); *Crimes (Sentencing) Procedure Act 1999* (NSW) s 23; *Criminal Law (Sentencing) Act 1988* (SA) s 10(1)(h); *Sentencing Act 1995* (NT) s 5(1)(h); *Crimes (Sentencing) Act 2005* (ACT) s 36. There are also similar provisions at the Commonwealth level. See *Crimes Act 1914* (Cth) s 16(2)(h).

¹³⁸ For an example of where a 50 per cent discount was allowed, see *R v Johnston* (2008) 186 A Crim R 345, 349–50 [15]–[21] (Nettle JA). For an application of these principles, see *Dan Ning Wang v The Queen* [2010] NSWCCA 319 (17 December 2010); *Yue Ma v The Queen* [2010] NSWCCA 320 (17 December 2010); *R v Nguyen* [2010] NSWCCA 331 (21 December 2010). This contrasts with the decision in *R v Sahari* (2007) 17 VR 269 where it was held undesirable to specify a specific discount for cooperating with authorities.

¹³⁹ One of the most fundamental rights, the right to a fair trial, is itself subject to a host of competing principles which delimit the scope and content of that right: see Mirko Bagaric, Theo Alexander and Marlene Ebejer, 'The Illusion That Is the Right to a Fair Trial in Australia' (2011) 17(2) *Australian Journal of Human Rights* 59.

who can make restitution. This is justified on the basis of the practical improvement to the lives of victims and the inherent logic in prioritising victim prosperity over abstract purity in sentencing.

*C Previous Good Character is Relevant: the Need for Doctrinal Coherency
in Light of the Progressive Loss of Mitigation Theory*

As discussed in part III of this paper, the prevailing approach to the relevance of previous good character in relation to white-collar offending is that it only marginally mitigates. The argument is that most white-collar offenders do not have prior convictions, and it is this very characteristic that enables them to secure status and positions that later provide the opportunity to offend. This analysis is flawed, at least in terms of consistency with the conventional treatment of prior offending, or lack of it, in the sentencing calculus.

Prior convictions have their most significant role in sentencing where offenders have a long criminal history, in which case they can lead to a considerably longer penalty.¹⁴⁰ This has fuelled criticism on the basis that it amounts to punishing the accused again for their previous crimes, that is, it constitutes double punishment and involves punishing people for their character, as opposed to what they have done.¹⁴¹

These criticisms have been met with the endorsement of the ‘progressive loss of mitigation’ theory, which is the view that a degree of mitigation should be accorded to first-time offenders or those with a minor criminal record. This mitigation is ‘used up’ by offenders who repeatedly come before the courts, thereby resulting in ever-increasing penalties for recidivists.

Hence, it is not that offenders with prior convictions are being punished more heavily, rather, that first-time offenders are treated more leniently. Recidivism disentitles repeat offenders to the leniency which is normally afforded the first-time offender.¹⁴²

In *Veen v The Queen [No 2]*¹⁴³ the High Court set out three other grounds for imposing harsher penalties on recidivists:

the antecedent criminal history is relevant ... to show whether the instant offence is an uncharacteristic aberration or whether the offender has manifested

¹⁴⁰ Approximately 60 per cent of imprisoned offenders have been to prison previously: see Australian Bureau of Statistics, ‘An Analysis of Repeat Imprisonment Trends in Australia’ (Research Paper No 1351.0.55.031, Australian Bureau of Statistics, August 2010) 12 <<http://www.ausstats.abs.gov.au/>>. Hence, this is a common aggravating factor.

¹⁴¹ Mirko Bagaric, ‘Double Punishment and Punishing Character: The Unfairness of Prior Convictions’ (2000) 19(1) *Criminal Justice Ethics* 10.

¹⁴² See, eg, *Baumer v The Queen* (1988) 166 CLR 51, 58.

¹⁴³ (1988) 164 CLR 465.

in his commission of the instant offence a continuing attitude of disobedience of the law. In the latter case, retribution, deterrence and protection of the society may all indicate that a more severe penalty is warranted. It is legitimate to take account of the antecedent criminal history when it *illuminates the moral culpability* of the offender in the instant case, or *shows his dangerous propensity* or shows a need to impose condign punishment to *deter the offender* and other offenders from committing further offences of a like kind.¹⁴⁴

Thus, it is well settled that, in general, good character results in less punishment. The fact that many offenders in a certain category of offending share this trait is no basis for diminishing its relevance. Rarity or commonality of a trait in relation to an offence type is not a basis for determining its role in the sentencing calculus. More than 90% of offenders plead guilty,¹⁴⁵ yet this does not diminish the weight given to the guilty plea discount. It follows that in relation to white-collar offending, good character should be accorded as much weight as in other types of offending.

D Extra-Curial Punishments Which Can Be Tangibly Measured Should Reduce Penalty

The other limb of the proportionality equation is the severity of penalty. Imprisonment is the harshest penalty in our system of law. However, a range of other deprivations can be imposed including partial loss of liberty or monetary exactions. For this article, the key issue is what types of losses should be regarded as relevant to this limb of the proportionality thesis: should this include only those losses directly imposed by the courts for sentencing purposes? Or should incidental hardships in the form of extra-curial hardship also be included?

As noted above, the main extra-curial penalty suffered by white-collar offenders is in the form of employment deprivation: either being dismissed from a job, being precluded from pursuing a certain career, such as a lawyer or accountant, or having those prospects severely curtailed as a result of a conviction. There is little doubt that these are damaging to the offender: ‘a person's employment is usually one of the most important things in his or her life. It gives not only a livelihood but an occupation, an identity and a sense of self-esteem’.¹⁴⁶

¹⁴⁴ Ibid 477 [14] (emphasis added). A similar approach was articulated in *Field v The Queen* [2011] NSWCCA 70 (21 April 2011). See also *Anderson v The Queen* [2010] NSWCCA 287 (7 December 2010); *Murphy v The Queen* (2005) 158 A Crim R 375; *Tsakonas v The Queen* (2009) 197 A Crim R 581; *Graham v The Queen* [2009] NSWCCA 212 (31 August 2009); *Wright v The Queen* [2008] NSWCCA 282 (15 December 2008); *R v Riddle* [2010] ACTCA 8 (25 May 2010); *Griffith v Tasmania* [2010] TASCRA 19 (11 November 2010) [14].

¹⁴⁵ Australian Bureau of Statistics, ‘Criminal Courts, Australia, 2009–2010’ (Statistics Publication No 4513.0, Australian Bureau of Statistics, 27 January 2011) <<http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/4513.02009-10>>.

¹⁴⁶ *Johnson v Unisys Ltd* [2003] 1 AC 518, 539 [35] (Lord Hoffmann).

The correct approach to factoring employment deprivations into the sentencing equation requires consideration of the very nature of punishment. Criminologists and philosophers generally adopt expansive definitions of punishment, recognising that incidental negative consequences flowing from a finding of guilt may constitute punishment. Antony Duff defines punishment as ‘the infliction of suffering on a member of the community who has broken its laws’.¹⁴⁷ Andrew von Hirsch states that ‘punishing someone consists of visiting a deprivation (hard treatment) on him, because he has committed a wrong, in a manner that expresses disapprobation on the person for his conduct’.¹⁴⁸ Similarly, punishment has been described as pain delivery.¹⁴⁹ Further, it has been asserted that ‘the intrinsic point of punishment is that it should hurt — that it should inflict suffering, hardship or burdens’.¹⁵⁰ Thus, drawing a line through the contemporary terminology, it would seem that punishment is a hardship or deprivation; the taking away of something of value for a wrong actually or perceived to have been committed. Notably, the definition does not take account of the forum (for example, a court, other institution or employer) which inflicts the hardship.

Thus it follows that although the experience of hardship by the offender in the form of employment deprivations may not be *intended* as punishment (but instead is, for example, to maintain the integrity of a profession) or may only be a *consequence* of punishment (for example, a statutory requirement), it does not mean that such deprivations do not of themselves constitute a punishment. As is evident from the definition above, what is crucial in this respect is the effect on the person, not the reason for which the hardship is inflicted. To ignore those hardships in the sentencing calculus is both illogical and contrary to the principle of proportionality.

The loss of a job or exclusion from an occupation tangibly sets back an individual's interests and there does not seem to be a justifiable reason for ignoring that aspect of a sentence on an offender. Thus, job status is a relevant criterion that should be recognised in the sentencing calculus. It has been argued that discrimination between offenders already exists in the form of (the privilege of) holding a job.¹⁵¹ To then allow that job to result in a sentencing discount only further compounds the disadvantage that is experienced by the unemployed offender. To this end, it has been argued that an offender who comes from a deprived social background should receive a sentencing discount because they have had less opportunity to lead

¹⁴⁷ R A Duff, *Trials and Punishments* (Cambridge University Press, 1985) 267.

¹⁴⁸ Andrew von Hirsch, *Past or Future Crimes: Deservedness and Dangerousness in the Sentencing of Criminals* (Rutgers University Press, 1985) 35.

¹⁴⁹ Nils Christie, *Limits to Pain* (Martin Robertson, 1981) 19, 48.

¹⁵⁰ R A Duff, ‘Punishment, Citizenship and Responsibility’ in Henry Tam (ed), *Punishment, Excuses and Moral Development* (Ashgate Publishing, 1996) 17, 18. A similar view was adopted by Jeremy Bentham who declared that ‘all punishment is mischief, all punishment in itself is evil’: Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation* (J H Burns and H L A Hart eds, Oxford University Press, 1970) 158.

¹⁵¹ See above, part III.

law-abiding lives and, hence, are less morally blameworthy. There is significant force in this argument. However, this is an argument which has not found favour with law-makers. Social deprivation is not, of itself, a mitigating factor because of the assumption that it would undermine the deterrent effect of criminal sanctions.¹⁵² Thus, the basic approach of sentencing law is that sentencing is not the place to redress social disparity and offenders must be taken as they are found at the time of sentencing. This is consistent with the approach to victims — they must be taken as they are, and offenders are liable for both the intended and the remote consequences of their conduct.¹⁵³

As noted in part III of this paper, there is some inconsistency and uncertainty regarding the impact of criminal wrongdoing on employment. This area of law is in need of fundamental reform. Employment deprivations are a form of punishment and, consequently, they should be recognised in the overall sanction meted out to an offender and subject to the normal sentencing practices and principles governing the infliction of criminal punishment.

This would also make sentencing law more consistent. In other instances, incidental deprivations stemming from offending are normally regarded as mitigatory. For example, a strong mitigating factor is incidental harm suffered by the offender, in the form of trauma associated with injuring or killing a friend or relative or also suffering serious injury.¹⁵⁴

Offenders who are not Australian citizens or permanent residents can be deported if they fail a ‘character test’; being sentenced to a term of imprisonment of a year or more can constitute evidence of bad character. Deportation or the risk of deportation is an additional burden that would then be faced by such an offender. Hence, logically it should be mitigatory. This was the position taken in *Valayamkandathil v The Queen*,¹⁵⁵ *Guden v The Queen*¹⁵⁶ and *Director of Public Prosecutions v Yildirim*.¹⁵⁷

Where an offender is harmed in the course of committing an offence, this can also reduce the penalty. In *Alameddine v The Queen*¹⁵⁸ the Court regarded the fact that the offender was injured when his drug-making laboratory exploded as a matter to be taken into account in mitigation.

To be clear, what should mitigate is not only the loss of a job as a result of a criminal sanction but also any employment deprivation. This extends to the diminished

¹⁵² Bagaric and Edney, above n 10.

¹⁵³ *DPP (Vic) v Eli* [2008] VSCA 209 (27 October 2008).

¹⁵⁴ See *R v Teh* (2003) 40 MVR 195.

¹⁵⁵ [2010] VSCA 260 (4 October 2010).

¹⁵⁶ (2010) 28 VR 288.

¹⁵⁷ [2011] VSCA 219 (28 July 2011).

¹⁵⁸ [2006] NSWCCA 317 (10 October 2006).

capacity of an offender to secure employment and the disqualification or suspension of a professional or similar qualification (such as in the areas of law, medicine or accounting) that often stems from a criminal sanction. These further hardships are normally less certain than the loss of a job and hence should carry less weight as penalty reductions than where it is clear that an offender will lose his or her job as a consequence of being found guilty of an offence. Moreover, the mitigatory impact of employment deprivations should apply even where the offence occurred in the employment setting.¹⁵⁹

The other form of incidental non-curial punishment is the censure that an offender may receive from family, friends, associates or the wider community. The extent of this sanction varies markedly according to such matters as the social connections and personal antecedents of the offender and the offence in question. This form of ‘punishment’ is too obscure and too hard to measure to be factored into the sentencing process. There is considerable logical force in the comments by Hayne J in *Ryan v The Queen*¹⁶⁰ where he stated:

There is an irreducible tension between the proposition that offending behaviour is worthy of punishment and condemnation according to its gravity, and the proposition that the offender is entitled to leniency on account of that condemnation.¹⁶¹

Although the public opprobrium and social ostracism suffered by an offender may be palpable to both the offender and the court, it should not factor into sentencing. However, this can only be a provisional view in the absence of a more extensive enquiry. It may be that bringing white-collar offenders publicly before the law constitutes both punishment and deterrence. Kahan and Posner suggest that, in certain circumstances, shaming might itself ‘provide sufficient deterrence to white-collar crime offenders’.¹⁶² But even if that is so, acting on such an assumption may result in an unjustifiable differential treatment between classes of offender.

E Curial Incidental Sanctions

The current legal position is that curial incidental sanctions are mitigatory. This position is correct and not in need of reform.

¹⁵⁹ This is to be contrasted with comments in *R v Liddy [No 2]* (2002) 84 SASR 231.

¹⁶⁰ (2001) 206 CLR 267.

¹⁶¹ *Ibid* 314 [157].

¹⁶² Dan Kahan and Eric Posner, ‘Shaming White-Collar Criminals: a Proposal for Reform of the Federal Sentencing Guidelines’ (1999) 42 *Journal of Law and Economics* 365, 368–72, cited in John B Owens, ‘Have We No Shame?: Thoughts on Shaming “White-Collar” Criminals, and the Federal Sentencing Guidelines’ (2001) 49 *American University Law Review* 1047, 1048.

*F Heavy Penalties Do Not Deter Other White-Collar Criminals:
Abolish the Emphasis on General Deterrence*

The strong emphasis by the courts on general deterrence when sentencing white-collar offenders is logically sound — especially in light of the assumption that such offenders are more likely to engage in a cost–benefit analysis than other offenders.

However, the assumption underlying this approach is not validated by the empirical evidence. The evidence on this front has been considered in a number of recent reviews. The studies show that harsh penalties, in the form of imprisonment generally and clearly disproportionate penalties, do not discourage crime.¹⁶³ The greatest deterrence against crime is not the size of the penalty, but the perceived likelihood of detection.¹⁶⁴ Of particular relevance is research establishing that this also applies to white-collar offences.¹⁶⁵ The upshot is that penalties for white-collar crime should not be increased in order to discourage other would-be offenders.

This does not mean that severe punishment is never suitable for white-collar offences. Rather, it means that a justification for severe sanctions must be located within other sentencing objectives. Moreover, if tough sanctions do not deter white-collar offences, there is a need for the criminal justice system to implement other measures which will achieve this outcome. It has been suggested that the answer rests in greater monitoring of areas where white-collar offences are often committed. Thus, in the context of tax crime, it might be that more audits should be undertaken. In other areas it might mean that greater measures to monitor compliance are implemented. More generally, a greater proportion of resources should be deployed in the areas of policing and crime detection, and even the pathology of white-collar offending warrants further investigation.¹⁶⁶

¹⁶³ See, eg, Raymond Paternoster, ‘How Much Do We Really Know About Criminal Deterrence?’ (2010) 100 *Journal of Criminal Law & Criminology* 765; Michael Tonry, ‘Mostly Unintended Effects of Mandatory Penalties: Two Centuries of Consistent Findings’ (2009) 38 *Crime and Justice* 65, 69; Cheryl Marie Webster, Anthony N Doob, and Franklin E Zimring, ‘Proposition 8 and Crime Rates in California: The Case of the Disappearing Deterrent’ (2006) 5 *Criminology & Public Policy* 417.

¹⁶⁴ Donald Ritchie, ‘Does Imprisonment Deter? A Review of the Evidence’ (Report, Sentencing Advisory Council, April 2011).

¹⁶⁵ Mirko Bagaric, Theo Alexander and Athula Pathinayake, ‘The Fallacy of General Deterrence and the Futility of Imprisoning Offenders for Tax Fraud’ (2011) 26 *Australian Tax Forum* 511.

¹⁶⁶ Marilyn Price and Donna M Norris, ‘White-Collar Crime: Corporate and Securities and Commodities Fraud’ (2009) 37 *Journal of the American Academy of Psychiatry and the Law* 538.

V CONCLUSION

Our analysis suggests that a doctrinally-sound approach to sentencing white-collar offenders requires a number of legislative reforms. The key reforms are:

- Providing a numerical and large discount for restitution — the argument that restitution allows offenders to buy their way out of gaol is outweighed by the good that comes from restoration for victims;
- Abolishing the pursuit of general deterrence and the increases in sentences which follow;
- Providing a greater discount for previous good character, consistent with other areas of sentencing;
- Providing a greater discount for non-curial harm in the form of employment deprivations; and
- Placing more emphasis on the harm caused by the offence to guide the sentence: crimes against individuals should be treated more seriously than those against the revenue or large corporations.

DAMAGES FOR BREACH OF CONTRACT: COMPENSATION, COST OF CURE AND VINDICATION

ABSTRACT

A cost of cure award seeks to undo the intangible harm or injustice that the promisor has caused to the promisee, thereby giving the promisee the performance he contracted for, and protecting his primary performance interest. In contrast, compensation protects the promisee against another kind of harm, although this time it is tangible: direct and/or consequential loss, which may flow from the violation of the promisee's primary right. The promisee's secondary compensation right requires the promisor to make good the promisee's pecuniary loss to undo the tangible harm that the promisee has also suffered from the breach. This article aims to illuminate the distinct role and aim of a cost of cure award as a legal response to breach of contract from compensatory damages, clarify when such a role should actually be fulfilled, and show how these two remedies can co-exist in harmony.

I INTRODUCTION

The assumption is often made that in order to be legally enforceable, a promise must, at least, entail a moral obligation. Samuel Stoljar says that the requirement to keep promises arises in the moral sphere before entering the legal domain.¹ When a promise is made, the promisee possesses a moral claim against the promisor, although the former may not yet have obtained the actual performance. In the words of Conrad Johnson, '[t]o be under a legal obligation is to be under a requirement that is ... a *moral* obligation'.² The legal obligation, therefore, represents the enforceable administration of morality. The moral obligation to perform the contract exists alongside the legal obligation.³

According to this explanation, the promisor is subject to two sets of obligations, namely legal and moral (where the former incorporates the latter). Therefore, when a judge, in the event of a breach, specifically enforces the promised performance, they are in fact doing no more than enforcing the promisor's moral obligation. The

* Assistant Professor of Law, School of Management, New York Institute of Technology.

¹ Samuel Stoljar, 'Keeping Promises: the Moral and Legal Obligation' (1988) 14 *Legal Studies* 258, 258.

² Conrad Johnson, 'Moral and Legal Obligation' (1975) 12 *Journal of Philosophy* 315, 332 (emphasis in original).

³ Nina C Z Khouri, 'Efficient Breach Theory in the Law of Contract: An Analysis' (2000) 9 *Auckland University Law Review* 739, 759.

legal duty is an extension of the moral obligation. These are the basic features of any legal regime that is in step with the morality of promissory obligations. If common law is to run parallel to the morality of promises, it must require the breaching seller to keep their promise, not simply to pay off the buyer.

The rule that expectation damages, not specific performance, is the primary remedy awarded for breach of contract is often taken to show that contract law does not recognise, and is not based upon, a moral obligation to perform a contract. This fact is at odds with the (moral) notion that ‘a contract is first of all a promise’ which, therefore, ‘must be kept because a promise must be kept’.⁴ However, a promissory conception of contract law does not require that specific performance be the default remedy. In spite of denying the non-breaching party specific performance, the common law of contract does not deny that the promise-breaker should have performed the contract and respected their moral obligation in the first place.⁵ Nor does it deny that specific performance is the ideal and most suitable moral response to promise-breaking or anticipatory repudiation. Yet there are convincing reasons why the courts should not always (through legal means) enforce the morality of promissory obligation or specific performance.⁶ Contract law, thus understood, can be said to run neither against the morality of promise nor parallel to it.

In common law specific performance is available only when damages are inadequate. A number of theorists have argued that the adequacy test is an unsatisfactory instrument for explaining why contract law, if it indeed holds that contractual obligation is based upon the moral obligation to keep one’s word, does not routinely grant specific performance in the event of a breach.⁷ This divergence between legal doctrine and the implication of a moral account of promising is so profound as to be in need of an explanation. If the traditional view that valid contractual obligations should be performed is indeed accepted in contract law, then why is specific performance not routinely granted in the event of promise-breaking?

According to the ordinary understanding of promissory obligation, judges should in general take seriously the moral force of contracts as promises. They should, in the event of a breach, always (through legal means) enforce the morality of promise or

⁴ Charles Fried, *Contract as Promise: A Theory of Contractual Obligation* (Harvard University Press, 1981) 17. See also Stephen A Smith, *Contract Theory* (Oxford University Press, 2004) 398; Dori Kimel, *From Promise to Contract: Towards a Liberal Theory of Contract* (Hart Publishing, 2003) 95–6.

⁵ Seana Shiffrin, ‘The Divergence of Contract and Promise’ (2007) 120 *Harvard Law Review* 708, 398–9; Kimel, above n 4, 104, 112.

⁶ Shiffrin, above n 5, 709, 733.

⁷ Kimel, above n 4, 96; Melvin Eisenberg, ‘Actual and Virtual Specific Performance, the Theory of Efficient Breach, and the Indifference Principle in Contract Law’ (2005) 93 *California Law Review* 975, 1016; Anthony Ogus, ‘Remedies’ in Donald Renshaw Harris and Denis Tallon (eds), *Contract Law Today: Anglo-French Comparisons* (Clarendon Press, 1989) 243, 243–63; Donald Harris, David Campbell and Roger Halson, *Remedies in Contract and Tort* (Butterworths, 2nd ed, 2002) 221–7; Andrew Burrows, *Remedies for Torts and Breach of Contract* (Oxford University Press, 3rd ed, 2004) 427.

specific performance. Once judges compel the defaulting promisor to provide the promised performance, his moral obligation becomes a legal obligation as well. The legal norms regulating promises must not diverge from the moral norms that apply to them; the legal duty must be an extension of the moral obligation. It therefore represents the enforceable administration of morality.

However, the failure to take seriously the moral force of contracts as promises is an indication that judges do not recognise a moral obligation to perform a promise, and thus run against the morality of promissory obligation. The moral obligation will not be satisfied if the defaulting promisor provides only a compensatory award.

The following questions then arise: Does the common law run against the morality of promise? The answer to this question is negative. A number of justifications have been put forward to explain the common law's reluctance to award specific performance despite its undoubted acceptance as the appropriate moral response to promise-breaking.

The justifications upon which we seek to justify the common law's approach are not driven by an underlying general normative position that contradicts the value of promissory norms as utilitarian theories do. For example, embracing any explicit or implicit recommendation to break a promise or asking the contracting parties to adopt the view that people who breach their contractual promises are doing a good thing if this would lead to better social consequences overall and that incentives should be produced for them to behave in this way.⁸ Judge Oliver Wendell Holmes argued that in law a contracting party has the option either to perform the contract (eg, sell goods, produce goods, render some service, and so on) or to pay damages for the loss.⁹ Therefore, if they provide compensation, they commit no wrong; they have done no more or less than to choose which way to perform their obligation. Holmes viewed the obligation to perform a contract as corresponding to a choice to perform or else to provide compensation.¹⁰

Holmes' view predates the so-called efficient breach theory. The idea underlying the theory is that the promisor should be allowed, or even encouraged, to break their contract and provide compensation instead if they can pursue a more profitable or desirable venture than that which they are currently pursuing, notwithstanding they have agreed to a contract and accepted its terms.¹¹ Lawyer-economists assume that

⁸ Shiffrin, above n 5, 732.

⁹ Oliver Wendell Holmes, 'The Path of the Law' (1897) 10 *Harvard Law Review* 457, 462.

¹⁰ Clark A Remington, 'International Interference with Contract and the Doctrine of Efficient Breach: Fine Tuning the Notion of the Contract Breacher as Wrongdoer' (1990) 47 *Buffalo Law Review* 645, 647.

¹¹ Maree C Chetwin and David K Round, 'Breach of Contract and the New Remedy of Account of Profits' (2002) 38 *ABACUS: A Journal of Accounting, Finance and Business Studies* 406, 410; Donald Harris, David Campbell and Roger Halson (eds), *Remedies in Contract and Tort* (Butterworths, 2nd ed, 2002) 12; R Sharpe, *Injunctions and Specific Performance* (Canada Law Book, 1983) 276.

an efficiently breached contract benefits the welfare of society, on the basis that the breach will result in a more efficient use of goods and services.¹² Therefore, undoubtedly influenced by Holmes' understanding of contractual obligation as a simple option either to perform or to provide compensation, lawyer-economists have analysed the law of contract as a means of determining and reinforcing efficient economic behaviour.¹³

The common law's reluctance to award specific performance, despite an undoubted acceptance that it is the appropriate moral response to promise-breaking, can be justified on many grounds. For example, it raises problems concerning both social norms and the risk of error,¹⁴ it creates a serious potential for opportunism,¹⁵ it is undesirable in a contract requiring personal services (eg, employment contracts) because this could be tantamount to slavery, and, in some cases, such as *Co-operative Insurance Society Ltd v Argyll Stores (Holdings) Ltd*,¹⁶ it is undesirable where performance requires undue judicial supervision.¹⁷

The present argument in this article is not intended to alter the common law's approach to specific performance, but rather to highlight that there are two distinct contractual interests constitutive of a contract and two distinct ways in which the promisor can cause harm to the promisee, each of which is protected by a different remedy and for a different purpose.

Contract law's concern appears focused on the legally binding contracts that market participants make and the rights and duties that occur as a result. A great deal of research has questioned the exact nature of these rights and duties.¹⁸ In particular, what rights does the buyer acquire by virtue of entering into a contract with the

¹² Frank Menetrez, 'Consequentialism, Promissory Obligation, and the Theory of Efficient Breach' (2000) 47 *University of California At Los Angeles Law Review* 859, 860.

¹³ Khouri, above n 3, 739.

¹⁴ Eisenberg, above n 7, 1020–1.

¹⁵ *Ibid* 1026.

¹⁶ [1998] AC 1. See also *Koepfel v Koepfel*, 161 NYS 2d 694 (App Div, 1957).

¹⁷ Dan B Dobbs, *Law of Remedies: Damages, Equity, Restitution* (West Publishing, 2nd ed 1993) 135–41. As Dobbs has pointed out:

The adequacy test is repeatedly invoked today when the plaintiff seeks equitable relief. Nevertheless its importance has declined. Many cases do not mention the test at all. Some cases mention the test but find plenty of grounds for saying that the legal remedy is not adequate. Adequacy of the legal remedy is often judged quite liberally in favor of the equitable remedy. When equitable relief is denied, it is quite often on grounds entirely distinct from the adequacy grounds ... So although the rule is invoked, it is also often ignored, sidestepped, or invoked in a way that means something else altogether. It is probably fair to say that the adequacy test has been evolving from a rule to a factor in the court's balance of costs and benefits ... The adequacy rule, as a rule that simply bars the gate, is virtually dead and probably should be (citations omitted).

¹⁸ See Rafal Zakrzewski, *Remedies Reclassified* (Oxford University Press, 2005); Stephen A Smith, 'The Law of Damages: Rules for Citizens or Rules for Courts?' in Djakhongir Saidov and Ralph Cunnington (eds), *Contract Damages: Domestic and International Perspectives* (Hart Publishing, 2008).

seller? What is each market participant obliged to do? In *Photo Production Ltd v Securicor Transport Ltd*,¹⁹ Lord Diplock made it clear that there are two kinds of rights: primary and secondary.²⁰ The distinction between primary and secondary rights, as far as English jurisprudence is concerned, was first made by John Austin.²¹ A secondary right arises from a legal wrong, a breach of a primary duty. A primary right is a right which exists independently of a wrong.

II PERFORMANCE AND COMPENSATION: THE TWO DISTINCT CONTRACTUAL INTERESTS

Private law protects many interests; for example, interests in a person's bodily integrity, and the entitlement to have contracts performed according to their specific terms (hereinafter referred to as 'the performance interest'). The imposition of legal duties on individuals requiring them to respect such interests, offers the primary method of protecting these interests.²² Individuals may be required to abstain from doing acts liable to harm these interests — for example, the duty to abstain from causing physical injury to another. Positive protection may also be given by a duty upon an individual to perform actions giving effect to the protected interest, for example, a duty to deliver goods or render services. Of course, these rights can be enforced directly so long as a claimant can obtain an injunction, requiring the defendant owing the duty to forbear from causing physical injury, or obtain a decree of specific performance, requiring the defendant owing the duty to do the very thing that the substantive duty requires them to do.²³ Pothier's original terminology²⁴ will be adapted for use in what follows: reference will be made to primary performance interests, which are protected by primary rights and duties.

It can thus be seen that enforcing a primary right directly involves ordering the defendant, who is under the corresponding duty, to do what they are required to do, or to forbear from doing what they are required not to do. Nevertheless, such enforcement will not always be possible. More specifically, it is not typically too late for the claimant to enforce (or rely on) their primary right when it is infringed. Nevertheless, it may be so if its subject matter is destroyed.²⁵ For example, a contractual right to deliver the Mona Lisa can no longer be given effect if the painting is destroyed. Similarly, a right to bodily integrity cannot be protected if the claimant

¹⁹ [1980] AC 827.

²⁰ *Ibid* 849.

²¹ Robert Joseph Pothier, *A Treatise on the Law of Obligations or Contract* (William David Evans trans, R H Small, 3rd ed, 1853). See also John Austin, *Lectures on Jurisprudence* (R Campbell (ed), John Murray, 5th ed, 1885) vol II, lecture XLV, 367.

²² Charlie Webb, 'Performance and Compensation: An Analysis of Contract Damages and Contractual Obligation' (2006) 26 *Oxford Journal of Legal Studies* 41, 42.

²³ *Ibid*.

²⁴ Pothier, above n 21, 367.

²⁵ Zakrzewski, above n 18, 104.

is killed by the defendant. It is futile to order the defendant subsequently to perform the original duty not to act in the way which caused the death of the claimant.

The same is equally true of a past infringement of a contractual right that correlates with a negative duty, in that it cannot be given effect to (or protected) by ordering the defendant not to breach their substantive primary duty.²⁶ So, there are two main and quite distinct contractual interests constitutive of a contract. First, the interest in securing the contracted-for performance; secondly, the interest in ensuring, if that performance is not completely (but substantially) secured or not secured at all, that one is not left worse off as a result thereof. The claimant can bring a claim to give effect to their performance interest and/or can bring a claim to give effect to their compensation interest. The secondary compensation interest is a separate interest, not merely an alternative or substitute formulation of the primary performance interest.²⁷

It may be argued, however, that in some cases, particularly where the claimant enters into the contract with a view to making a profit, although they obtain damages rather than performance, they end up in an identical position, having effectively made their

²⁶ Ibid 105.

²⁷ The well-known English case of *Farley v Skinner* [2002] 2 AC 732 illustrates the difference between a primary obligation and a secondary obligation, between a performance interest claim and a compensation interest claim. In that case, Farley bought a house that was situated 15 miles from Gatwick Airport. He hoped the house was a quiet country retreat and, before the purchase, he appointed the defendant to survey the property and specifically asked him to investigate and report whether the property was affected by aircraft noise. The defendant stated that noise was unlikely to be a problem, ‘although some planes will inevitably cross the area, depending on the direction of the wind and positioning of the flight path’: at [750C]. On the strength of the surveyor’s negligent assurance that the house was not significantly affected, the claimant decided to purchase it. However, after spending a considerable sum in refurbishment, the claimant discovered that aircraft noise was quite often a problem and that it interfered with his enjoyment of the house. Although the claimant’s enjoyment of the property was diminished by the noise, the value of the house was nevertheless unaffected. The claimant decided not to sell the house but sued the surveyor for damages for discomfort and inconvenience arising out of the breach of contract. The trial judge found that the defendant was in breach of contract and awarded the claimant £10 000 for physical inconvenience, discomfort and distress. The House of Lords upheld this award. Inasmuch as the surveyor’s defective performance was less valuable than the performance due, the difference between these two values was paid by the surveyor to Farley as compensatory damages (£10 000). Surely, the combination of the defective performance and the compensatory award places Farley in no worse a position than that which he would have been in if the contract had been performed properly. However, he has not received the contracted-for performance — that is, of a surveyor exercising a reasonable care to investigate — with the consequence that Farley will not be in the position of having received the performance due to him. This means that Farley’s performance interest is not given effect. His right to the contracted-for performance is not being enforced. Here, the court has recognised a separate interest and right in Farley, namely, his compensation interest in not being left worse off by reason of the surveyor’s breach of contract. As Webb notes, ‘[t]here is a difference between recovering the value placed on the stipulated performance and actually receiving that stipulated performance’: Webb, above n 22, 55. See also Smith, above n 18, 36–7.

profit (ie, the expected increase in their wealth is still fulfilled). Should receiving compensatory damages for non-performance in such cases be understood to be the same as receiving performance? This question must be answered negatively. Consider the following example. Suppose that a seller agrees to sell 200 bicycles to a buyer for a total price of \$1000. But suppose further that, before delivery, a third party (ie, a second buyer), who values the seller's bicycles more highly than the first buyer, offers the seller \$2500 to purchase the 200 bicycles. Suppose the seller cannot meet the third party's demand without breaching their contract with the first buyer, as they have only 200 bicycles to sell. If the seller breaches their contract with the first buyer, they must pay them damages. The market price of the 200 bicycles at the time of breach is \$1200, and the first buyer's valuation of the bicycles at the time of breach is \$1400. If the contract had been performed, the first buyer would have enjoyed a surplus of \$400. The first buyer will enjoy a surplus of \$400 if they recover market-price damages, because they can purchase an identical 200 bicycles on the market for \$1200, and their market-price damages of \$200 will bring their effective price to \$1000.

It is true that the first buyer, after the receipt of the damages (\$200), will be able to make the profit they intended, but careful consideration shows us that before they are able to recover the \$200 which will eventually protect their reason for entering into the contract (ie, they will be able to make the profit they expected), they have to go into the market and purchase 200 bicycles that are identical to the contracted-for bicycles. They must then sue for the difference between the contract price and the market price. This tells us that the first buyer's market-price damages of \$200 are designed merely to ensure that their financial loss is made good; that they are not left worse off by reason of the seller's breach of their substantive primary duty to perform. The \$200 only remedies the losses that flow from the violation of the buyer's primary right to performance. The secondary obligation, thus understood, does not entitle the first buyer to the performance due to them (ie, 200 bicycles).

In contrast, the aim of the primary obligation is different. Instead of seeking to compensate the first buyer for any loss caused by the seller's breach of contract, it seeks to ensure that they do receive the consignment of bicycles for which they bargained — the primary obligation remedies the breach itself. The important point is that, although compensatory damages might in some cases mean that the claimant achieves and fulfils the main reason for entry into the contract, a claim for compensation for the claimant's financial loss still does not amount to an assertion of their performance interest. It should therefore be distinguished from a performance interest claim. Compensatory damages merely respond to the financial loss which may flow from the breach of the claimant's primary right to performance and not to the breach itself.²⁸ This distinction can be best explained by the theory of corrective justice.

²⁸ Catherine Mitchell, 'Remedial Inadequacy in Contract and the Role of Restitutionary Damages' (1999) 15 *Journal of Contract Law* 133, 150; Catherine Mitchell, 'Promise, Performance and Damages for Breach of Contract' (2003) *Journal of Obligations and Remedies* 67, 69.

III CORRECTIVE JUSTICE

The idea of corrective justice received an early formulation in Aristotle's treatment of justice in *Nicomachean Ethics*, Book V.²⁹ For Aristotle, corrective justice is the theory of the mean; more specifically, 'the just, or the equal, is the mean between the more and the less'.³⁰ Once it is established that the defendant has, as a result of their wrongful act, taken and acquired more than they ought to have — that is, more than the mean — then they must surrender their surplus to the claimant, who has less than the mean, or who has less than what they would have had, had the defendant never acted wrongfully towards them.³¹ As a result of the wrong there is an excess (gain) for the defendant, while the claimant endures deficit (loss) as a result of an injustice at the defendant's hands.

That Aristotle refers to the gains and losses of corrective justice normatively, rather than materially or financially, is indisputable. He considers that the equality between the particular parties is disturbed whenever corrective justice is violated. In this way, he lays the complete normative weight of his theory on that equality. The question now to be examined is this: in what regard could the parties possibly be equal? Aristotle provides no clear answer to this crucial question. He simply offers corrective justice as a transactional equality, without saying in what respect the parties are equal. The result is that we cannot, in a dialogue, merely state that the defendant's behaviour is an 'injustice', because merely to state this does not provide an argument. We must explain why the word 'injustice' arises, or is applied in the first place, which requires an account of the kind of equality applicable here, and an account of why it is wrong to disturb it without justification.

The theory of corrective justice is a philosophical explanation — first outlined by Aristotle and later allegedly incorporated by Immanuel Kant into the notion of natural right — of how justice may be done in private law for both parties.

The Kantian principle of right 'is a philosophy of freedom that starts with the operation of free will conceived as self-determining activity'.³² In Kant's account, '[t]he fundamental principle applicable to the interaction of self-determining beings is that action should be consistent [or co-exist] with the freedom of whomever the action might affect'.³³ According to Kant, rights — such as contractual performance — 'are the juridical manifestations of the freedom inherent in self-determining

²⁹ Aristotle, *Nicomachean Ethics* (Roger Crisp trans, Cambridge University Press, 2000) Book V, 2–5, 1130–3, [a14]–[b28].

³⁰ Francesco Giglio, 'Restitution for Wrongs: A Structural Analysis' (2007) 20 *Canadian Journal of Law and Jurisprudence* 5, 22.

³¹ *Ibid.*

³² Ernest J Weinrib, 'The Gains and Losses of Corrective Justice' (1994) *Duke Law Journal* 277, 290. See Immanuel Kant, *The Metaphysics of Morals* (Mary J Gregor trans, Cambridge University Press, 1991) 40–3.

³³ Weinrib, above n 32, 290–1.

activity'.³⁴ Action is thus compatible with the freedom of others so long as it is not contrary to their legal rights. If the promisee has a right to contractual performance, the other party is morally bound by a corresponding obligation to perform unless the promisee has released them from that obligation. The promisee has control over the choices available to the promisor who bears the corresponding obligation.³⁵ To put this differently, the promisee is in a moral position to determine, by their freedom of choice, the way in which the promisor should behave and in this way to limit the latter's freedom of choice.³⁶ In the words of Kant, rights are 'moral capacities for putting others under obligations'.³⁷

Right and obligation are connected — and so therefore are the promisor and the promisee — by the fact that the substance of the right is the essence of the obligation. The right represents the moral position of the promisee, which is to ensure that they demand and receive just what they have been promised in the contract; the promisee cannot demand more than that. The obligation represents the moral position of the promisor, which is to ensure that they perform no less than what they consented to perform in the contract. If both positions are maintained as stipulated, then the promisor and promisee are on an equal footing with respect to their rights and obligations in the contract. The promisor's freedom of action should be capable of coexisting with the freedom of the promisee, which manifests itself in the right to performance, always assuming that the two freedoms must coexist, with the two sides being equal.³⁸ As Weinrib states:

As Aristotle himself notes, the parties to a corrective justice transaction are equal in a very peculiar way: the equality abstracts from the particularity of the parties' social rank or moral character to the sheer relationship of wrongdoer and sufferer. Corrective justice treats the parties as equals because all self-determining beings, regardless of rank or character, have equal moral status. The conjunction of right and duty is simply this equality of self-determining beings viewed juridically, from the standpoint of the correlativity of one person's action and its effects on another.³⁹

³⁴ Ibid 291; Kant, above n 32, 90–5, 101–3.

³⁵ Eduardo Rivera-López, 'Promises, Expectation, and Rights' (2006) 81 *Chicago-Kent Law Review* 21, 34.

³⁶ H L A Hart, 'Are There Any Natural Rights?' (1955) 64 *The Philosophical Review* 175, 180; Peter Vallentyne, 'Natural Rights and Two Conceptions of Promising' (2006) 81 *Chicago-Kent Law Review* 9, 12.

³⁷ Kant, above n 32, 63.

³⁸ See Stoljar, above n 1, 269.

³⁹ Weinrib, above n 32, 292 (citations omitted). For a more complete argument of the connection between corrective justice and Kantian right, see Peter Benson, 'The Basis of Corrective Justice and Its Relation to Distributive Justice' (1992) 77 *Iowa Law Review* 515, 601–24; Steven J Heyman, 'Aristotle on Political Justice' (1992) 77 *Iowa Law Review* 851, 860–3; E Weinrib, 'Restitutionary Damages as Corrective Justice' (2000) 1 *Theoretical Inquires in Law* 1.

This statement can be best explained through the following example. Suppose that an employer enters into a contract with their employee prohibiting them from selling and disclosing any confidential information during their term of employment and thereafter. The employee has an obligation, which means that there is something owed specifically to the employer, so a legal right arises out of this contract. If the employee breaches their contractual promise to the employer by selling and disclosing confidential information to a third party, the employer will claim that the breach represents a wrong against them — that wrong arising from the claim that they have been unequally, unjustly or harmfully treated in the sense of diminishing their status as a promisee. The employer's moral status to determine, by their freedom of choice, how the employee should use the information (and in this way to limit the latter's freedom of choice to act) has been diminished. Thus understood, the absence of coexistence between these two freedoms would simply mean that the employee will cause normative (or as I call it 'intangible') harm to the employer. In this light, the employee's breach of contract leaves the employer in a normatively disadvantaged situation. The two parties are no longer equal in their moral status.

Unless the employee can undo the wrong or injustice they have committed (by breaching the primary right not to sell and disclose confidential information to a third party), the employer will never be able to re-establish their condition as controller of how the information ought to be used by the employee, and, thereby, as a promisee, with respect to the past infringement of the contractual right. However, it is a foregone conclusion that the wrong committed by the employee cannot be undone. The employer cannot require the employee to refrain from doing what they have already done — 'the past cannot be undone'.⁴⁰ As Zakrzewski said, '[r]equiring the person owing the duty to abstain from doing what he or she has already done would be a fruitless exercise'.⁴¹ Therefore, regarding this past infringement, the employer cannot regain their status as a promisee. The freedom of both parties can in no way be returned to a state of coexistence. The state of equality, which involves the employee and the employer being on an equal footing with respect to their rights and obligations in the contract, can no longer be achieved or restored.

The employer can, it is true, rely on this primary right to regulate future conduct by obtaining an injunction to prevent the employee from committing any further infringements, but this primary right will provide no protection with respect to the infringements already committed. Here, the employer can bring a claim for compensatory damages to make good their pecuniary loss concerning the past infringements — the secondary compensation interest.

Clearly, Ernest Weinrib's work provides some valuable clues in what respect the parties are equal. His normative approach provides a background to the idea of equality. In the following section, I attempt to situate the concepts of intangible and tangible harm within the theoretical framework of corrective justice supplied by Weinrib.

⁴⁰ Zakrzewski, above n 18, 105.

⁴¹ Ibid.

IV INTANGIBLE AND TANGIBLE HARMS

If the promisor breaches their obligation to perform — and thus the promisee's right to performance — they have then acted unequally, unjustly or harmfully against the promisee. For example, in the sense of undermining their position as a promisee, the promisee's moral status to determine, by their choice, how the promisor should act has been diminished (hereinafter referred to as 'intangible harm'). The concept of intangible harm it can, therefore, be seen, arises in the absence of the coexistence of the freedoms of, or equality between, the two parties, regardless of whether or not the promisor's breach of promise has actually caused the promisee any financial loss. Stated differently, the promisor's breach of contract is intangibly harmful to the promisee, not because it necessarily deprives them of a financial interest (although it sometimes, perhaps often, does so), but because it leaves them in a disadvantaged situation: their situation as a promisee has been undermined.⁴² The intangible harm here is thus independent of any material or financial measurement. It is a normative concept, which refers to the disadvantageous position occupied by the promisee as a result of the promisor's breach of contract.

By compelling the breaching promisor to fulfil their duty to perform, and so the promisee's right to performance, the court seeks to undo the intangible harm or injustice that the promisor has caused to the promisee. The court also restores to the promisee the privilege of limiting the promisor's freedom of choice of how to act, which was undermined by the promisor's behaviour, thereby giving the promisee the performance they contracted for, and protecting their performance interest. The promisee's interest in having the promise performed is a primary interest, which is effectuated by the recognition of the promisee's (primary) right that the promisor should perform their side of the contract. This brings about a corresponding (primary) obligation on the promisor to perform. Correctly understood, the performance interest does not seek to prevent or remedy the financial loss that the promisee may suffer by reason of the promisor's breach of the primary duty. After all, there are cases where, although the promisor's performance was defective, the promisee suffered no financial loss.⁴³ The promisee's claim for compensation could not therefore be linked with their performance interest claim. The performance interest in such cases

⁴² See Giglio, above n 30, 25.

⁴³ Consider the following example. Suppose that a contractor promises to build a house to certain specifications, one of which is that Brand X pipes are to be used in the plumbing. The contractor builds the house according to the specifications, save that they used different materials, installing Brand Y pipes rather than Brand X. In order to calculate the claimant's financial loss from this breach, the court must determine what the claimant stood to gain from the performance of the contract. Inasmuch as Brand Y is equal in quality, appearance, market value and cost to Brand X, the use of Brand Y pipes does not affect the value of the building work (whether this is assessed at market rates or by reference to the value placed on the work by the claimant). Accordingly, no financial loss is suffered by the claimant. But still the claimant has not received the exact performance they contracted for. In such a case, therefore, if the claimant aims to force the defendant to deliver the promised performance, it will be difficult to argue that compensation can give effect to their interest in having the contract performed as

seeks to make the promisor perform what they have promised, no more, no less. This primary interest is protected if a prohibitory injunction or a specific performance remedy is available to the promisee. As Charlie Webb has agreed:

Where a claimant asserts his performance interest the notion of loss is superfluous. He has a right that the defendant performs and he is seeking to have this enforced. It may be that a breach of contract will cause him loss ... but his interest in performance is independent of this. A contracting party has a right to the defendant's performance regardless of what losses, if any, will be caused by failure to perform.⁴⁴

There are, therefore, two distinct ways in which the promisor can cause harm to the promisee, each of which is protected in a different way and for a different purpose. The promisee has a secondary compensation interest in not being left worse off by reason of the promisor's breach of primary duty. This interest is effectuated by recognising a right in the promisee that the promisor should compensate the promisee for any financial losses resulting from failure to perform their primary duty.⁴⁵ It protects the promisee against another kind of harm, although this time it is tangible: direct and/or consequential loss, which may flow from the violation of the promisee's primary right (hereafter referred to as 'tangible harm').

The promisee's secondary compensation right requires the promisor to make good the promisee's pecuniary loss to undo the tangible harm that the promisee has also suffered from the breach. For, when the time has passed for the delivery to be made on time, only the promisor's breach of the first part of the obligation (viz, for delivery to be made) and, therefore, the intangible harm suffered by the promisee in the sense of them being undermined in the first instance, has been undone by specific performance. Here, the promisee's primary performance interest is substantially fulfilled, but they have still suffered financial loss for services having been delivered late. The promisor's breach of the second part of the obligation (viz, that delivery be made on time) and, as such, the intangible harm suffered by the promisee in the sense of them being undermined in the second instance, cannot be cured and undone, where the time has already passed for the delivery to be made on time. This being so, it follows that the promisor is required to protect the promisee's secondary compensation interest, which ensures that the promisee is not left worse off as a result of not having had their primary performance interest completely addressed. The promisee will then be entitled to be awarded the amount of their pecuniary loss, let's say \$15 000 here, as the monetary value calculated to equal the value of timely delivery to the promisee. The secondary interest, thus understood, does not seek to undo the intangible harm or injustice that the promisor has caused to the promisee, and so

⁴³ *cont'd*

specified. This indeed proves that compensatory damages cannot and should not be said to equate to enforced performance. This example is based on the facts of *Jacob & Youngs, Inc v Kent*, 230 NY 239 (1921).

⁴⁴ Webb, above n 22, 54.

⁴⁵ See Zakrzewski, above n 18, 102–3, 165–6; Smith, above n 18, 36–7.

does not give the promisee the performance they contracted for. Rather, the remedy of compensation, unlike specific performance prohibitory injunction, responds to loss resulting from the breach and not to the breach itself.

V SMITH'S INTANGIBLE HARM CONCEPT

Stephen A Smith argues that if a promisor has committed a breach of contract, then they will cause both a tangible harm (ie, direct and/or consequential loss) to the promisee and an intangible harm.⁴⁶ It should be noted here that by 'intangible harm' Smith does not mean the injustice that the promisor has committed against the promisee in terms of his diminished status as a promisee, as described earlier. Nor does he mean the loss of reputation, injured feelings, pain and suffering, or indeed any kind of 'mental distress', 'emotional disturbance' or 'aggravated damages' that may be cited as grounds for compensation as a result of a breach.⁴⁷ Such losses are still regarded as tangible harms.⁴⁸ For Smith, the meaning of intangible harm is understood to be the destruction of the actual or potential bonds of trust between the parties to the contract.⁴⁹

In Smith's view, an award of damages is a wholly effective means to compensate the promisee for the tangible harm caused to them as a result of the promisor's failure to perform, but it is not an adequate means of repairing the intangible harm.⁵⁰ Compensation, he says, cannot remedy the intangible harm that a breach has caused, because bonds of trust are created by the promisor's voluntary undertakings. It is not possible to buy them.⁵¹ The trust that the promisee has put in the promisor is lost upon breach and it cannot be restored by monetary damages. Can one buy friendship or love? Of course not; and the same conclusion applies to broken bonds of trust in promise cases.

One might wonder, then, whether a decree of specific performance could remedy the intangible harm caused by a breach in any better way. The answer is an unequivocal 'no'. According to Smith, to compel the promisor to provide the promised performance following breach of the contract is 'self-defeating' given that the only way to give effect to the bonds of trust created by the promisor's voluntary undertakings is through performance which itself is voluntary.⁵² Where there is coercion, trust is necessarily not manifest. The relationship of mutual trust and respect, which

⁴⁶ Stephen A Smith, 'Performance, Punishment and the Nature of Contractual Obligation' (1997) 60 *Modern Law Review* 360, 396.

⁴⁷ See Guenter H Treitel, *The Law of Contract* (Sweet and Maxwell, 12th ed, 2007) 891–4.

⁴⁸ Smith, above n 46, 396.

⁴⁹ *Ibid.*

⁵⁰ *Ibid.* 370.

⁵¹ *Ibid.*

⁵² *Ibid.*

has broken down, cannot be rebuilt by an order of specific performance. Accordingly, compelling the promisor to perform amounts to recognition not of the special bond that connects them to the promisee, but rather of the existence of external sanctions.⁵³ An order of specific performance is a highly coercive remedy. Violating such an order constitutes contempt of court and is punishable by jail, a civil fine or both. These background prospects are the main reason why the defaulting promisor performs their undertaking in accordance with such an order. Bonds of trust between contracting parties are created by agreement and cannot be created by an order of specific performance. Such an order, thus understood, is an ineffective means to remedy the intangible harm resulting from the promisor's failure to perform the contract.

Smith concedes that valid contractual undertakings should be performed. But, whereas there is no remedy for intangible harm, Smith argues that adopting a regime of routine specific performance 'would make the intangible benefits of contracting more difficult to achieve'.⁵⁴ In his view, this is the case because potential breachers would perform their obligations in order to escape the threat of the award of a decree of specific performance and its consequences if not complied with. If so, then it follows that adopting a regime of routine specific performance would leave the interacting parties ignorant of 'whether or not performance has been done for the right reasons' and even make it less likely that it has.⁵⁵ Each no longer knows whether the commitment of the other to fulfilling their obligations results from them keeping their word or merely from a desire to escape the decree and its sanctions.

It further follows, from Smith's view, that a regime of routine specific performance would deprive performance, and consequently contracting in general, of its potential to create bonds, thereby 'making the valuable activity of contracting less valuable'.⁵⁶ In practical terms, the bond-creating function of contract would be weakened in the case where specific performance is the normal legal response to a breach of contract. The exclusion of specific performance as the default rule would conversely make the promisor's performance of the contract a stronger confirmation of the special bond between them and the promisee.⁵⁷ Thus understood, if contract is to be taken seriously as a way of creating bonds, then the recognition of specific performance as a general rule must be rejected. The justification for routinely dealing with breach of contract by an award of monetary damages in preference to specific performance is that more scope is thereby left 'for the operation of good intentions'.⁵⁸ As Smith says:

There are some wrongs that law cannot undo and some aspects of life into which law should not intrude. The bond-destruction that results from breach

⁵³ Ibid.

⁵⁴ Ibid.

⁵⁵ Ibid 370–1.

⁵⁶ Ibid 371.

⁵⁷ Ibid.

⁵⁸ Ibid.

of contract is a wrong that law cannot cure and the bond-creation that results from the completion of a contract is an area of life into which law should not intrude.⁵⁹

Smith's argument, however, is not persuasive. More specifically, he assumes that when the promisor breaches their promise, the promisee files a specific performance suit against them only because they have destroyed the relationship which had been or was being developed between them. This is not necessarily so. For, although it is true that the promisee normally has a complaint concerning the promisor's abuse of their trust, and that there is no remedial response to this sort of complaint, we should also not forget that the promisee has another complaint. It is that the promisor has acted unequally, unjustly or harmfully towards them in the sense of diminishing their status as a promisee. Thus, the promisee seeks specific performance to undo that injustice, thereby obtaining the performance contracted for. This paper does not question the fact that the promisee's complaint regarding the destruction of the special tie between them and the promisor has no remedial response. Rather, it simply holds that Smith has overlooked the fact that the promisee has another complaint against the promisor, one which can still be responded to through the remedy of specific performance. Smith's intangible harm concept is wrongly defined. He should instead adopt my definition, namely, that the promisor's breach of contract is intangibly harmful to the promisee, because it leaves them in a disadvantaged situation: their status as a promisee has been undermined.

VI THE PERFORMANCE INTEREST AND DAMAGES AWARDS

The performance interest is given effect by particular remedies: specific performance and injunctions. Nevertheless, the question to be considered now is how the performance interest may be given effect by an award of damages.

To begin with, the court may award the claimant damages measured on a 'cost of cure' basis. On this basis, the claimant is given the sum of money needed to obtain the performance for which they contracted from a different source. Consider the following example. Suppose that farmer Macdonald agrees to allow the Gritty Gravel Co. to extract gravel from some of his fields. The contract requires Gritty to restore the fields to their previous condition when extraction is complete. Gritty refuses to do so. Following this breach, Macdonald asks the court to enforce the contract. Rather than providing Macdonald with a decree of specific performance, however, the court may award him damages measured on the cost of cure. Such an award will enable Macdonald to obtain the stipulated performance (ie, restoring the fields to their previous condition) from another source.

The gravel extraction example shows that a cost of cure award is not an ordinary damages award. Typically, compensatory damages merely respond to any direct and/or consequential loss which may flow from the breach of the claimant's primary

⁵⁹ Ibid 377.

right to performance and not to the breach itself. It is clear that compensation must be compensation for loss. Yet a cost of cure award is intended to allow claimants such as Macdonald to cure the breach itself. It aims to place him in the actual position he would have been in had Gritty performed its primary duty; it does not compensate for any financial losses. Value of loss damages and cost of cure damages do not share the same objective — they are fundamentally different. For example, in *Radford v de Froberville* (*'Radford'*),⁶⁰ the defendant breached his obligation to build a wall that would separate his land from the claimant's. The claimant sued for the cost of building the wall. The court awarded the cost of cure although the difference between the value of the claimant's land with or without the wall was almost nil. The cost of cure is clearly greater than the value of loss, and therefore cannot be understood as compensatory. It is rather a substitute for ordinary specific relief. The damages awarded in *Radford* are a substitute for, and a vindication of, the right that has been infringed by the defendant. As Smith has argued:

[C]ost of cure awards are best explained as a form of substitute specific relief: their aim is not to compensate the plaintiff for the value or utility of whatever was lost, but to eliminate or undo the physical change in the plaintiff's world that has been or will be brought about by the defendant's breach of duty ... Payment of the cost of cure is a substitute for what the defendant should have done originally. Having failed in his primary substantive duty (to perform a contract, to not injure), there arises, at the moment of failure, a substitute duty to achieve the same end by paying for substitute performance.⁶¹

However, both the gravel extraction example and *Radford* show that a cost of cure award provides the claimant only with the resources to obtain from a different source the stipulated performance they contracted for. Does this change the fact that such an award is equivalent to performance of the contract — and that therefore the

⁶⁰ [1977] 1 WLR 1262.

⁶¹ Smith, above n 18, 35–6. For a similar view on this matter, see Charlie Webb, 'Justifying Damages' in Jason W Neyers, Richard Bronaugh and Stephen G A Pitel (eds), *Exploring Contract Law* (Hart Publishing, 2009) ch 6. Charlie Webb observes that if the right to performance is taken seriously, then it would seem that specific performance should be the natural remedy and most suitable moral response to promise-breaking. However, the occasion for the performance may be past or its reclamation may require undue supervision by the court. If this is so, then what is the solution? Webb claims that this contractual right 'can, sometimes, be effectuated through an award ... which the claimant uses to purchase an equivalent performance from an alternative source'. The aim is to find remedies which in so far as possible abide by the special domain of contract law. Webb argues that any damages award that goes beyond simulating performance requires reflection on 'the norms and ideals which shape and justify the law': at xiv. For more details on the 'cost of cure' award, see Stephen A Smith, 'Substitutionary Damages' in C Rickett (ed), *Justifying Remedies in Private Law* (Hart Publishing, 2008); Ewan McKendrick, *Contract Law* (Oxford University Press, 6th ed, 2005) 1017; Ewan McKendrick, 'Breach of Contract and the Meaning of Loss' (1999) 52 *Current Legal Problems* 37; Brian Coote, 'Contract Damages, *Ruxley*, and the Performance Interest' (1997) 56 *Cambridge Law Journal* 537; Eisenberg, above n 7, 1041–8.

claimant's performance interest is effectively met? The answer is 'no'. Although the end product will be obtained from another source, a cost of cure award gives the claimant a substantial part of the performance, thereby making only a small change to the nature of their performance interest. After all, as Guenter H Treitel has said:

Specific relief will hardly ever give the aggrieved party exactly the performance to which he was entitled, if for no other reason than that the enforced performance resulting from this form of relief will generally take place at a time other than that stipulated in the contract.⁶²

In practice, the claimant contracts both for services to be delivered and for those services to be delivered on time. Should the court enforce only one of these obligations, it is not providing the claimant with all they contracted for. If the court is, however, content in such circumstances to order a different performance from that agreed upon between the parties, on the basis that the claimant thereby gains a substantial part of the performance originally agreed upon, then there appears to be no objection to the court doing so by making a cost of cure award. The claimant would thus obtain the end product for which they contracted. In both cases it is difficult in the end to give the claimant exactly what they contracted for, but still it cannot be denied that they receive a substantial part of what the stipulated performance would have given them.⁶³ There is a great similarity between specific performance and a cost of cure award. Each gives the claimant a substantial part of the performance they contracted for; they can be an excellent means of protecting the claimant's primary performance interest.

One reason why the court might be reluctant to grant the claimant a specific performance decree and instead grant them a cost of cure award can be found in the fact that the defendant's performance of their original obligation is not possible anymore, because it is physically impossible for them.⁶⁴ Furthermore, the court's reluctance to compel the defendant specifically to perform can be justified because the threat of enforcing specific performance could be used oppressively — for example, where the defendant, in order to circumvent a difficult personal confrontation, might be required to pay the cost of cure, buying out the claimant's right to specific performance.⁶⁵ It can also be justified based on the risk that because of the great tension between the two parties, the defendant who is required to perform specifically may not fulfil their obligations properly and carefully, but may instead perform defectively. A cost of cure award is therefore a substitute for specific performance, awarded in cases where specific performance is not possible or desirable. It is a vindication of the right that has been infringed by the defendant.

⁶² Guenter H Treitel, *Remedies for Breach of Contract – A Comparative Account* (Oxford University Press, 1988) 1.

⁶³ Webb, above n 22, 61.

⁶⁴ The fact that the actions specified in a contract are physically capable of being performed does not mean that the defendant is still able to perform their primary contractual duty. See Smith, above n 18, 36.

⁶⁵ Harris, Campbell and Halson, above n 7, 191.

One crucial question remains: are there any conditions that should be satisfied in order for the claimant to be entitled to the cost of cure award? The answer is an unequivocal ‘yes’. There are two such conditions. First, the claimant is entitled to the cost of cure award if they undertake to use it to obtain the end product of performance. More specifically, as noted above, the aim of the cost of cure award is not to compensate the claimant for any financial losses caused by the defendant’s failure to deliver the stipulated performance — ie, the notion of loss is simply superfluous — but rather to enable the claimant to obtain that stipulated performance from a different source, thereby protecting their primary performance interest. Thus, such an award aims to place the claimant in the actual, not just metaphorical, situation in which they would have been if the breach had never been committed in the first place; it is a substitute for what the defendant should have done originally.

Surely, if such an award is not used for this purpose, the claimant will end up with defective performance and a sum of money which they can spend as they wish, with the consequence that the cost of cure award cannot be said to be equivalent to performance of the contract. The claimant must not be allowed to keep such a windfall. This is because if a cost of cure award is justified on the ground that it is a substitute for specific performance for the protection of the performance interest, then it seems as a matter of reasonableness that such an award must be given to the claimant, if, and only if, they undertake actually to use it to do the relevant work.⁶⁶ Therefore, clearly, it will be unreasonable to grant the claimant the cost of cure award, if they have no intention to cure the breach (or to effect reinstatement) — the rationale for the award disappears. Oliver J concluded in *Radford*, discussed above, that ‘the claimant genuinely wants this work done and that he intends to expend any damages awarded on carrying it out. In my judgment, therefore, the damages ought to be measured by the cost of the work, unless there are some other considerations which point to a different measure’.⁶⁷ The money should not simply go straight into the claimant’s pocket as a windfall but should instead be spent on curing the breach and therefore reflects the claimant’s interest in securing the contracted-for performance; in seeing the contract they are involved in being performed. Otherwise, such an award cannot be regarded as giving effect to the claimant’s performance interest, and therefore cannot be justified on this basis. As Webb has said:

[T]he reason why the claimant should not be entitled to such a windfall is not because this exceeds any loss that he may have suffered (it may, but this is not point where the claim is not for compensation), but because this in no way gives effect to his performance interest. This being so, there is every reason why the court *should* be concerned with the use to which the claimant may put a performance interest damages award ... such an award should be conditional upon the claimant undertaking to use his damages.⁶⁸

⁶⁶ Smith, above n 4, 422.

⁶⁷ [1977] 1 WLR 1262, 1284.

⁶⁸ Webb, above n 22, 63 (emphasis in original) (citations omitted).

While one important reason for the existence of a concept of ‘reasonableness’, above, is to avoid the conferral of a windfall on the claimant, it can also be seen to have a role beyond this, namely, to avoid undue hardship to the defendant. This is actually the second condition that should be satisfied in order for the claimant to be entitled to the cost of cure award, namely, that granting the claimant such an award does not cause severe hardship to the defendant.⁶⁹ The example of pipes, discussed earlier,⁷⁰ will elucidate this. In that example, the house as built conforms to the specifications except in relation to the brand of pipes to be used in the plumbing work during the construction. Brand Y pipes rather than Brand X were installed. However, Brand Y is equal in quality, appearance, market value and cost to Brand X — indeed, the same thing, though manufactured in another place. The pipe was largely encased within the walls of the completed house. In light of this, to conform to the requirements of the owner would involve knocking down the house and rebuilding. Here, the claimant should not be entitled to the cost of cure award, for they have already received substantially what they had contracted for. It would be wholly unreasonable to grant the claimant such an award ‘in order to enable him to undo what had been done when what had been done was substantially what he had contracted for’.⁷¹ This is an example where it would clearly be unreasonable to award cost of cure.

The fact that the claimant has already received substantially what they had contracted for means that granting them such an award would cause undue or severe hardship to the defendant.⁷² The defendant has a claim to being treated fairly, although it is them who breached the contract and the claimant who is the victim.⁷³ However, the problem in the pipes example is that if the proper measure of damages is the difference in value and the diminution in value is nil, then should the court not revert to an award on the cost of cure? The answer is ‘no’. Lord Lloyd, in *Ruxley Electronics and Construction Ltd v Forsyth* (‘*Ruxley*’), considered that an injustice that arises by way of making an award of too little is not counterbalanced by a court making an award of too much: ‘that cannot make reasonable what ... has been found

⁶⁹ Similarly, specific performance will not be ordered where this would cause undue hardship to the defendant (see, eg, *Patel v Ali* [1984] Ch 283).

⁷⁰ Above n 43.

⁷¹ Ewan McKendrick, ‘Breach of Contract and the Meaning of Loss’ (1999) 52 *Current Legal Problems* 37, 48.

⁷² See Stephen A Smith, ‘The Law of Damages: Rules for Citizens or Rules for Courts?’ in Djahongir Saidov and Ralph Cunnington (eds), *Contract Damages: Domestic and International Perspectives* (Hart Publishing, 2008) 35–6; Edward Yorio, ‘In Defense of Money Damages for Breach of Contract’ (1982) 82 *Columbia Law Review* 1365, 1389; Webb, above n 5, 63–4; Guenter H Treitel, *The Law of Contract* (Sweet and Maxwell, 11th ed, 2003) 947–8, 1026–7 (discussing how the cost of cure award can be refused on the ground of severe hardship to the defendant, as where, for example, the expense of the work involved is wholly out of all proportion to the benefit to be obtained).

⁷³ See Edward Yorio, *Contract Enforcement: Specific Performance and Injunctions* (Little, Brown, 1989) 527; Paul Richards, *Law of Contract* (Pearson Education, 8th ed, 2007) 399–400.

to be unreasonable'.⁷⁴ So, compelling the defendant in the pipes example to pay the cost of cure would cause a hardship amounting to injustice. Where the cost of cure award is unreasonable, the claimant will be compensated, for example, for 'loss of amenity' or mental distress. This is a middle position between two extremes and it is fair to both parties. The result might be otherwise if the claimant had not received substantially that for which they had contracted; for example, where a house is constructed so defectively that it is of no use for its designed purpose. In such a case, it would be reasonable to award cost of cure to undo the breach (or to effect reinstatement). This being so, the question of reasonableness plays an important role in deciding whether to make the award for the cost of cure or not.

In *Ruxley*, neither of these two conditions were satisfied. In that case, the claimants entered into a contract with the defendant to construct a swimming pool in his garden. It was to be of a maximum depth of seven feet six inches, but in breach of contract, the finished pool was only six feet nine inches deep. When the defendant discovered this fact, he refused to pay the contract price. The claimants filed a suit against him for the contract price. The trial judge found that the claimants had substantially performed their obligations under the contract. In this light, it was held that the defendant was liable to pay the claimants the contract price. The defendant was thus left to his counterclaim for damages for breach of contract.

At trial, three facts were established. The first was that the difference in value between the swimming pool which was provided and the pool which should have been provided was found to be nil. The second was that the only way to increase the depth was to rebuild the pool at a cost of £21 460, which was nearly a third of the total price of the pool. The third was that the difference in depth did not impair the defendant's use of the pool. The trial judge, in light of the above facts, refused to award the defendant the cost of cure of £21 560. Rather, he awarded him damages to reflect the loss of amenity which he suffered as a result of not getting the pool for which he had contracted. The Court of Appeal (by a majority) allowed an appeal by the defendant, holding that the decision to award the defendant general damages of £2500 for loss of amenity did not adequately protect his interest in having the contract performed, because the award would not enable him to have the swimming pool for which he contracted.

The course followed by the majority of the Court of Appeal was thus to award the defendant the full cost of carrying out the work. The House of Lords overturned the decision of the Court of Appeal and refused to award the defendant the cost of cure of £21 560, as to do so would be unreasonable.⁷⁵ The cost of cure was described by Lord Mustill as 'wholly disproportionate'⁷⁶ to the value of the loss and by Lord Lloyd as 'out of all proportion'.⁷⁷ Moreover, their Lordships did not

⁷⁴ [1996] AC 344, 374.

⁷⁵ In *Radford and Ruxley*, in deciding whether the cost of cure was an appropriate award, the courts referred to whether such an award was reasonable.

⁷⁶ [1996] AC 344, 354.

⁷⁷ *Ibid* 357. See also *McGlenn v Waltham Contractors Ltd* (2008) 111 Con LR 1.

actually believe that the defendant was as interested in getting his pool rebuilt as he was in punishing the claimants. Their Lordships were therefore concerned that the defendant was seeking the cost of cure award in bad faith. However, rather than awarding the defendant no damages at all for breach of contract the House of Lords upheld the trial judge's ruling for loss of amenity (£2500).

In light of the above discussion, it seems that there are two main and quite distinct contractual interests constitutive of a contract: the performance interest and the compensation interest. The secondary interest is a separate interest, not merely an alternative formulation of the primary interest: receiving compensatory damages for non-performance is not the same as receiving performance. However, some scholars entirely reject this distinction. For example, in his book *From Promise to Contract*, Dori Kimel talks about wrongful losses as the only harm that the promisee can possibly suffer as a result of the violation of their right to performance.⁷⁸ He believes that protection against a wrongful loss (ie, a loss that arises from rights-infringing behaviour) can be provided in two ways. The first is by preventing the promisor from causing wrongful (financial) loss by ordering them to do the very thing they had promised to do under the contract. The second is by redressing (or repairing) the wrongful loss by ordering the promisor to compensate the promisee for the financial loss flowing from the violation of their right to performance. In Kimel's view, the court gives the promise-breaker the choice to perform the contract or to pay damages in order to prevent or repair wrongful losses when — and only when — damages would fully (or effectively) redress those wrongful losses.

According to this view, Kimel seems to perceive both compensatory damages and specific performance as protecting the promisee against one kind of harm and so giving effect to one interest, namely, the performance interest. In other words, he seems to perceive the performance interest as the sole contractual interest and reimbursement orders as substitutional remedies which protect (or give effect to) the performance interest, thereby leaving no place for a secondary (compensatory) interest. Obviously, then, the distinction between the performance interest and the compensation interest has been neglected by Kimel. For him, thus understood, receiving compensatory damages for non-performance is the same as receiving performance. In what follows, this article questions Kimel's rejection of the distinction. More precisely, this article suggests that there are at least two reasons why this view is unsustainable and why Kimel's view should therefore not be embodied in contract law.

A Freedom of Contract

A simple explanation of why Kimel's understanding of contractual obligations should not be embodied in the law — and accordingly why we should accept that there are two distinct contractual rights and interests — can be derived from the principle of freedom of contract. More specifically, Kimel's view that the court gives the promise-breaker the choice to perform the contract or to pay damages in order

⁷⁸ Kimel, above n 4, 96–104.

to prevent or repair wrongful losses does not reflect what the contracting parties agree to. Certainly, parties could agree to disjunctive obligations to perform or to compensate; indeed, some parties clearly do reach such agreements. However, as a general account of how parties understand their agreements, it is quite unconvincing to suggest that they have agreed, in the typical case, to disjunctive obligations to perform or to provide compensation.⁷⁹ In practice, parties enter into a contract in the expectation that it will be performed⁸⁰ — a conclusion supported by empirical evidence. In 1990, David Baumer and Patricia Marschall surveyed 119 North Carolina corporations about their attitudes towards wilful breach. One question was, ‘[i]f a trading partner deliberately breaches a contract because a better deal can be had elsewhere, is such behaviour unethical?’ One hundred and five respondents said ‘yes’.⁸¹ Contracting parties do not usually accept that performance may depend on a subsequent choice by the other party to pay damages rather than perform.⁸² Rather, it seems likely that if this contingency were explicitly considered by the parties, they would decide to grant a right to terminate the contract upon payment of damages, or otherwise provide for the eventuality of breach.⁸³ As Lord Lindley said in *South Wales Miners’ Federation v Glamorgan Coal Company Ltd*:

Any party to a contract can break it if he chooses; but in point of law he is not entitled to break it even on offering to pay damages. If he wants to entitle himself to do that he must stipulate for a provision to that effect.⁸⁴

If Kimel’s view were applied by judges, they would examine the actual content of a contractual obligation and determine whether enforcing it would be economically efficient. To put this differently, instead of requiring judges to look into the procedural

⁷⁹ See Smith, above n 4, 400, 402.

⁸⁰ Khouri, above n 3, 756.

⁸¹ David Baumer and Patricia Marschall, ‘Willful Breach of Contract for the Sale of Goods: Can the Bane of Business be an Economic Bonanza?’ (1992) 65 *Temple Law Review* 159, 165. In addition, 86 participants said that they would always or almost withhold future business from a party who had wilfully breached: at 166.

⁸² This view is also supported in Ian Ayres and Gregory Klass, ‘Promissory Fraud Without Breach’ (2004) *Wisconsin Law Review* 507, 513–4 (citations omitted) (emphasis in original), arguing that:

[T]here are good reasons why promisors want to implicitly say that they intend to perform simpliciter, rather than that they intend to perform or pay damages, or that they do not intend not to perform, or nothing at all about their intent. Promisees care about promisor intent because they care deeply about whether or not the promisor will perform. If a promisee thinks that the promisor does not intend to perform and is seriously considering the option of paying damages instead, he is much less likely to rely on her promise, be it by entering into a binding contract or by otherwise ordering his behavior as if performance were going to happen. But the whole point of promising is to convince others to rely on one’s future actions. Thus promisors have a natural incentive to communicate with their promisees an intent to perform. This fact explains why most promises represent an intent to perform and why the law should adopt a default interpretation that recognizes this fact’.

⁸³ Khouri, above n 3, 756.

⁸⁴ [1905] AC 239, 253.

form of the contract, Kimel's view requires them to look into the actual content of each contract so as to decide the existence and scope of contractual obligation based on the concept of economic efficiency.⁸⁵ If it is possible to breach the contract efficiently, then such a breach will be allowed and encouraged. The promisee's right to performance will not be recognised and only their right to compensation will be protected. The promisor will be allowed to revisit the decision they previously made by intentionally placing himself under an obligation to perform — subject, of course, to the payment of damages. The incorporating of Kimel's view into the law would mean that the judges would not adequately respect the intention of the parties involved — that is, what they have actually said or written in their contract — and their freedom to determine their rights and obligations.⁸⁶ The traditional and still orthodox notion of the nature of contractual obligations is that they are voluntarily assumed.⁸⁷ Their content is thus a matter for the parties who have voluntarily and freely entered into the contract, not the state.⁸⁸ The principle of freedom of contract rejects the interference of the state in the terms of contracts. As Sir George Jessel MR said:

[I]f there is one thing which more than another public policy requires, it is that men of full age and competent understanding shall have the utmost liberty of contracting, and that their contracts when entered into freely and voluntarily shall be held sacred and shall be enforced by Courts of Justice.⁸⁹

It is clear, therefore, that Kimel's view is inconsistent with the idea of freedom of contract. It may be argued, however, that the court may often supplement or place some limitations on the obligations expressly agreed upon between the parties. In view of this, there are already restrictions upon freedom of contract. That being so, then the fact that Kimel's view of contractual obligations consists of another restriction upon — or undermining of — freedom of contract is insignificant. If there was nothing wrong with other pre-existing restrictions upon freedom of contract or with its undermining, then what can be wrong in this instance?

The answer is that Kimel's view introduces intervention despite the fact that the express terms of the contract leave clear how parties understand their agreement and how the dispute or problem between them should be resolved. Thus, the intervention cannot be reasonably justified in this instance. Take the case of *Harvela Investments*

⁸⁵ Khouri, above n 3, 757.

⁸⁶ Ibid; Daniel Friedmann, 'The Efficient Breach Fallacy' (1989) 19 *Journal of Legal Studies* 1, 23.

⁸⁷ A 'voluntarily assumed obligation' is created by an action which indicates that one intends to place oneself under an obligation to do something. See Neil MacCormick and Joseph Raz, 'Voluntary Obligations and Normative Powers', (1972) *Proceedings of the Aristotelian Society*, Supplementary Volumes 46; Joseph Raz, 'Promises and Obligations' in Peter Hacker and Joseph Raz (eds), *Law, Morality and Society* (Clarendon Press, 1977) 210.

⁸⁸ See Hugh Collins, *The Law of Contract* (Butterworths, 4th ed, 2003) 25; Smith, above n 4, 59; Laurence Koffman and Elizabeth Macdonald, *The Law of Contract* (Tolley, 4th ed, 2004) 3–4.

⁸⁹ *Printing and Numerical Registering Co v Sampson* (1875) LR 19 Eq 462, 465.

Ltd v Royal Trust Co of Canada Ltd.⁹⁰ In that case, the first defendants invited the claimants and the second defendant to submit a single sealed offer for the shares and stated that they would accept the highest complying offer. The first defendants were effectively offering a simple contractual arrangement whereby the shares would be sold to the highest bidder amongst those who submitted offers. The claimants tendered a fixed bid of \$2175 000 and the second defendant tendered a referential bid of \$2 100 000 or \$101 000 in excess of any other offer, whichever was the higher. The first defendants accepted the second defendant's bid, treating it as a bid of \$2 276 000, but the Court held that the first defendants were obliged to accept the claimants' bid.

The referential nature of the second defendant's bid rendered it invalid, as it was not within the terms of the contract to bid (because it was not a single offer). Obviously, there is no lacuna in the provisions of the contract, because it expressly provides for the situation which has occurred. In this light, the first defendants were contractually bound to sell the shares to the claimants. Kimel's view, however, introduces intervention by allowing the first defendants to breach the contract with the claimants and enter into a contract with the second defendant (subject, of course, to the payment of compensatory damages), despite the fact that the terms of the contract are clear and offer excellent guidance.

It must be realised that the courts do not usually interfere with the obligations expressly agreed between the parties unless faced with situations not dealt with by the express terms of the contract. For example, John contracts to sell his car to Mike, who offers \$2500 to purchase it, but for reasons unforeseen by both parties, the car is totally destroyed an hour after the conclusion of the agreement. What is the legal position in cases of this type? How should the dispute be resolved? There is clearly a lacuna in the provisions of the contract, because it does not expressly provide for the situation which has occurred. In such cases, the courts step in to fill the gaps left by the parties in sensible and reasonable ways, using a legal rule, the doctrine of frustration of contract, which dictates that in the event of certain unexpected contingencies the court may consider the contract as being terminated.⁹¹ In this view, the contractual obligations of both parties are automatically terminated and neither is obliged to perform. Thus, neither can be held liable for breaching the frustrated contract.⁹²

⁹⁰ [1986] AC 207.

⁹¹ See, for example, *Javad v Aqil* [1991] 1 WLR 1007.

⁹² Collins, above n 88, 298–9; Koffman and Macdonald, above n 88, 434–5. See, eg, *National Carriers v Panalpina (Northern)* [1981] AC 675; *Davis Contractors Ltd v Farebam UDC* [1956] AC 696, 698. The statutory law of sales cancels further performance of the contract if the goods perish without fault of either party before ownership has passed to the buyer (see, eg, *Sale of Goods Act 1979* (UK) c 54 s 7; United States *Uniform Commercial Code* ('UCC') § 2-615. Intervention is also introduced to protect various types of relatively weak or vulnerable contract parties, for example, consumers (see, eg, *Unfair Contract Terms Act 1977* (UK) c 50), employees (see, eg, *Employment Protection (Consolidation) Act 1978* (UK) c 44; *Employment Act 1989* (UK) c 38; *Trade Union and Labour Relations (Consolidation) Act 1992* (UK) c 52, s 37(1)(a)), tenants (see, eg, *Rent Act 1977* (UK) c 42; *Landlord and Tenant Act 1985* (UK) c 70)). For further examples, see Kimel, above 4, 118–9.

Accepting that parties should enjoy the freedom to choose the terms of their own contract entails an acceptance that, in all contracts, the duty of the defendant to perform (which is correlative to the claimant's right that the defendant perform) should not be regarded as equivalent to an option to either perform or pay damages. This in turn requires the recognition of performance and compensation as two distinct contractual rights and therefore interests.

B The Notion of Breach

A further simple explanation of why Kimel's view should not be embodied in the law — and therefore why we should accept that there are two distinct contractual rights and interests — is found in the notion of breach of contract. More specifically, in the ordinary understanding, a breach of contract is a breach of the duty to perform the original (primary) obligation (eg, deliver goods, produce goods, perform a service, and so on). Thus, its premise is that the defendant has a duty to perform. If no duty exists on the part of the defendant to perform their (primary) original obligation, then there can be no such thing as a breach of contract.⁹³ Breach, thus understood, presumes a duty to perform, because in the absence of such a duty there is plainly nothing that can be breached. That the defendant has an obligation to perform and that the claimant, correspondingly, has a right to the performance of the defendant is what establishes the primary interest of the claimant in performance.⁹⁴ It gives effect to that interest.

Failure to realise the distinctiveness of the claimant's performance interest leads to the position where the defendant's failure to perform would not in itself constitute a breach of duty. Thus, a claim for compensation would be in the nature of a claim to enforce a primary interest instead of a claim with regard to a secondary interest established upon a breach of duty.⁹⁵ Primary rights are distinct from the secondary rights which arise upon breach of duty. They give effect to different interests. That English contract law incorporates the notion of breach of contract and recognises distinct rights and interests in performance and in compensation is clear: judges do not typically describe a contractual obligation as an obligation to either perform or pay damages in lieu of performance. Rather, they maintain that there is an obligation to perform a contract and describe its breach as a wrong.⁹⁶ As Oliver LJ stated in *George Mitchell (Chesterhall) Ltd v Finney Lock Seeds Ltd*: 'the purpose of a contract is performance and not the grant of an option to pay damages'.⁹⁷

Nevertheless, Kimel's view gives the contracting parties the choice either to perform the contract or to pay damages for losses resulting from breach. It is clear that for Kimel compensatory orders enforce the original (primary) obligation. Both specific

⁹³ See Webb, above n 22, 46.

⁹⁴ *Ibid.*

⁹⁵ *Ibid.* 49.

⁹⁶ Smith, above n 4, 119.

⁹⁷ [1983] QB 284, 304.

performance and damages are regarded as forms of *true* specific performance. In this view, there is no difference between performance and compensation, between primary and secondary obligations. Kimel perceives compensatory damages and specific performance as giving effect to the same interest: the performance interest. In practical terms, the theory perceives this as the sole contractual interest and reimbursement orders as substitutional remedies which give effect to the performance interest, thus leaving no place for a secondary (compensatory) interest. Adopting Kimel's view that a claim for compensation is in the nature of a claim to enforce a primary interest, instead of a claim with regard to a secondary distinct interest established upon a breach of duty, results in an absurdity. The result is that we must deny that a contract imposes a duty to perform on the part of the defendant and that the contract may in fact be breached. The notion of breach of contract has no place in Kimel's view — it is a fiction.

So, in the example of the employee who has breached their contract with the employer by selling and disclosing confidential information, it would be erroneous to view the secondary right as a substitute method of protecting or giving effect to the primary interest. The employer's right to receive damages for the financial losses resulting from the breach of primary duty does not protect their interest that the employee not sell and disclose the confidential information. The same conclusion applies in respect of the primary right in the sense that it would be erroneous to view both the primary and secondary rights as protecting a single interest of the employer — that the employee should ensure that the employer's financial loss is made good. In fact, to view them in this way would result in paying 'insufficient attention to the existence of the primary right and duty and the fact that causing [such financial loss] is treated as a breach of duty by the [employee]'.⁹⁸

One must realise that the employee's primary duty is not to infringe the employer's protected primary interest, and accordingly not to cause them any financial loss. It should not be understood as a duty simply to pay monetary damages in order to reimburse the employer if the employee causes them such financial loss. If it were, then the act of selling and disclosing the confidential information which caused the financial loss would not be seen or perceived as a breach of duty. A conflation of the primary and secondary interests in the sense of understanding the employer's interest as being in the protection of the confidential information or, should that information be disclosed by the employee, in ensuring that the employee's behaviour causes the employer no loss, leaves no scope for any notion of breach of contract.⁹⁹ As a matter of principle, therefore, accepting that all contracts may be breached entails an acceptance that all contracts impose a duty to perform. This in turn requires the recognition of performance as an interest distinct from the compensation interest.

⁹⁸ See Webb, above n 22, 43.

⁹⁹ Ibid.

VII CONCLUSION

To sum up, it can be stated that there are two distinct contractual interests and two distinct ways in which the promisor can cause harm to the promisee, each of which is protected by a different remedy and for a different purpose. First, the promisee has an interest in seeing the contract they are involved in being performed. If a promisor breaches this interest, then they cause intangible harm to the promisee. The performance interest seeks to restore to the promisee the privilege of limiting the promisor's freedom of choice of how to act, which was undermined by their (the latter's) behaviour, thereby giving the promisee the performance they contracted for. It neither prevents nor repairs the financial loss that the promisee has suffered due to the promisor's breach. This interest is protected if a prohibitory injunction, a specific performance remedy, or a cost of cure award is available to the promisee. Second, the promisee has an interest in not being left worse off by reason of the promisor's breach of primary duty to perform. The compensation interest protects the promisee against another kind of harm, namely, tangible harm: the financial loss which flows from the violation of the promisee's primary right to performance. Understood in this way, the secondary interest is a separate interest, not merely an alternative formulation of the primary interest: receiving compensatory damages for non-performance is not the same as receiving the performance itself.

Since there are these two distinct contractual interests and, therefore, two distinct ways in which the promisor can cause harm to the promisee, Kimel must be taken to have adopted an artificially narrow concept of harm in order to secure his desired conclusion. By talking about wrongful losses, as if it is the only harm that the promisee can possibly suffer as a result of the violation of their right to performance, Kimel assumes that there is just one contractual interest. The danger here is that when the promisee brings a claim to give effect to their primary right to performance, it will be understood to be the same as a claim to repair losses caused by the promisor's wrongful breach and that it is, therefore, this right to which compensation for the financial loss flowing from the violation will give effect. However, the claim is really there to undo the intangible harm that the promisor has caused to the promisee.

TRANSSEXUALISM AND THE CONSIDERATION OF SOCIAL FACTORS WITHIN SEX IDENTIFICATION LAW

ABSTRACT

The law relating to sex identification has undergone a period of rapid change across the last forty years. Facets of contemporary Australian law have abandoned the use of biologically-focused tests in favour of a new multifactorial approach that considers multiple biological, psychological and social factors when determining the legal sex identity of a transsexual person. Although this course of development has increased legal recognition of the psychological sex identification of transsexuals, this article argues that legal consideration of social factors has the capacity to harden into a restrictive vector of normative control. This article warns that such consideration is potentially problematic because it may undermine the importance of a transsexual person's subjective experience of their sex identity, bring extensive and intrusive legal interrogatory pressure to bear on their life-histories and social lives, and require them to conform to narrow, stereotypical models of sex before legal recognition is awarded.

I INTRODUCTION

In October 2011, the High Court of Australia's decision in *AB & AH v Western Australia*¹ granted the applications made by two transmen² to be issued with recognition certificates that legally identify their sex as male. Judicial consideration was given to the fact that both applicants had been on hormone therapy for years and had their breasts surgically removed, and the Court placed emphasis on the social recognition of the applicants' sex as being male. Notably, however, neither applicant had undergone surgical sex reassignment of their genitalia. This decision stands in stark contrast to one made forty years earlier in *Corbett v Corbett*,³ in

* Assistant Professor, School of Law, University of Western Australia. I would like to acknowledge and thank the University of Western Australia for the financial assistance my research has received through the Baillieu Research Scholarship. I would also like to thank the anonymous reviewer for the *Adelaide Law Review* for their insightful and helpful comments on an earlier draft of this article.

¹ (2011) 244 CLR 390.

² Transsexuals who are born biologically female but who self-identify and live as male are known as 'transmen'; similarly transsexuals who are born biologically male but who self-identify and live as female are known as 'transwomen'.

³ [1971] P 83.

which Ormrod J emphasised strict biological criteria relating to genitalia, gonads and chromosomes as the only relevant considerations for identifying a person's legal sex identity. Whereas in 1971 Ormrod J decided that a transwoman remained legally male even though she had replaced her male genitalia with surgically constructed female genitalia, in 2011 the High Court decided that the transmen before the Court were legally male even though they not only lacked surgically constructed male genitalia but also retained female genitalia.

Clearly, the law relating to sex identification has altered significantly in just forty years. Underpinning this change is the ongoing broadening of the focus on biological factors inherited from Ormrod J's conceptualisation of legal sex identity as existing strictly in biological terms. However, this article is more concerned with where the development of sex identification law is heading, rather than where it has come from. As Australian sex identification law moved away from the narrow focus on biology, it initially shifted towards a model encompassing consideration of both biology and psychology. Now, however, some areas of the law have shifted again, towards a broad test encompassing consideration of multiple factors including biology, psychology, social recognition, lifestyle and personal history. This development appears *prima facie* to chart a progressive course away from the reductive application of narrow rules to transsexual bodies, and towards the holistic legal analysis of complete transsexual subjects. However, legal change in this area should not be read uncritically as a simple 'opening up' of the law that in all ways increases the access of transsexuals to legal recognition of their psychological sex identity. Rather, this article argues that although the once largely insurmountable biological-based legal barriers have been abandoned, judicial consideration of social factors could erect new legal barriers to recognition because it has the capacity to operate in restrictive and normative ways.

This article works through this argument in two parts. Part II tracks the development of Australian law around sex identification for transsexuals, demonstrating how it has moved from a narrow focus on biological factors to a wider focus on multiple factors, including social factors. In doing so, Part II addresses the key cases and statutes that have marked this development across the last forty years, and specifically focuses on the increasing emphasis on social factors in these sources of law. Part III unpacks this article's core critique about the consideration of social factors within sex identification law. It argues that even though these factors are less strict than the traditional biological factors, they still have the capacity to diminish the access of transsexuals to legal recognition of their psychological sex identity and to lead to inequitable outcomes. It identifies and engages with three specific areas of concern, warning that legal consideration of social factors could lead to the undermining of the importance of transsexuals' own psychological sex identifications and subjective experiences, the bringing of extensive and intrusive legal interrogatory pressure to bear on their lives, and the requirement that they conform to narrow, stereotypical models of sex.

II DEVELOPMENTS IN SEX IDENTIFICATION LAW

The law regarding the legal recognition of the psychological sex identification of transsexuals is bound up with developments in medical technology. Although

medical efforts to ‘change’ the sex of human patients began from around the early 1900s,⁴ the development of effective antiseptics, less dangerous and more effective anaesthetics, and more complex medical treatments and techniques, meant that sex reassignment procedures became more widely available from the 1950s onwards.⁵ The exact types of procedures that individual patients choose to undergo varies, but may include the amputation of sex-specific born biological structures (such as through mastectomy, penectomy and vaginectomy), the surgical construction of new sex-specific biological structures (such as through breast implantation, phalloplasty and labiaplasty), and the sex-specific aestheticisation of appearance (such as through facial feminisation surgery). In conjunction with these surgical options, regular hormone dosages (ie, oestrogen for transwomen and testosterone for transmen) are typically used in the medical treatment of transsexuals, and can significantly alter their bodies in terms of hair growth, fat distribution, and genitalia size and function.⁶

A Strict Biological Criteria

*Corbett v Corbett*⁷ is the first major case in the Western common law tradition that Australian law has drawn upon to decide whether or not these sex ‘change’ medical treatments affect a transsexual person’s legal sex identity. Ormrod J was required to determine whether April Ashley Corbett, a transwoman, could be considered a ‘woman’ for the purpose of marriage law (wherein marriage was defined as only being possible between a ‘man’ and a ‘woman’). April had been born biologically male but had received hormone therapy for a number of years, lived and socialised as a woman, had both her penis and testicles removed and had a vagina surgically constructed in their place. In the course of deciding that April’s marriage was invalid because both April and her husband were legally ‘male’, Ormrod J declared that the criteria for falling within the definition of ‘woman’ ‘must ... be biological’ in nature.⁸ He determined that the law should utilise three biological criteria to

⁴ Joanne Meyerowitz, ‘Sex Change and the Popular Press; Historical Notes on Transsexuality in the United States, 1930–1955’ (1998) 4 *GLQ: A Journal of Lesbian and Gay Studies* 159, 161.

⁵ Vern L Bullough, ‘Legitimizing Transsexualism’ (2007) 10 *International Journal of Transgenderism* 3, 4. It is important to flag a terminological issue here. Although widely referred to as ‘sex reassignment surgery’, surgical procedures intended to alter the biological sex characteristics of transsexual patients have also recently begun to be referred to as ‘sex affirmation surgery’: see, eg, Australian Human Rights Commission, *Sex Files: The Legal Recognition of Sex in Documents and Government Records* (2009). Although the reasoning behind this terminological shift is laudable — it recognises the persistence and centrality of a transsexual person’s psychological sex identification over and above their born biology — I have not adopted it in this article because the law treats such surgeries as important factors to be considered when it comes to *changing* the legal sex identification of transsexuals.

⁶ See, eg, the list of changes hormone treatment made to the transmen applicants in *Western Australia v AH* (2010) 41 WAR 431, 437 [7], 438 [11].

⁷ [1971] P 83.

⁸ *Corbett v Corbett* [1971] P 83, 106.

decide sex: chromosomal, gonadal and genital.⁹ If congruency is found between these three factors then Ormrod J argued that this ‘determine[s] the sex for the purpose of marriage accordingly’.¹⁰ The fact that April no longer had male genitalia was irrelevant, these criteria were to be judged in relation to born biology and thus April’s subsequent sex reassignment surgery did ‘not affect her true sex’.¹¹

As the first detailed and comprehensive engagement of modern UK law with the issue of sex identification, *Corbett v Corbett*¹² was destined to be an important precedent. It was cited favourably and applied in *R v Tan*,¹³ a criminal law case wherein it was necessary to determine whether a transwoman was a ‘man’ under statute law relating to prostitution. In applying the strict biological criteria from *Corbett v Corbett*,¹⁴ the Court rejected ‘without hesitation’ the submission that if a person ‘had become philosophically or psychologically or socially female, that person should be held not to be a man’ under the law.¹⁵ Despite the fact that *Corbett v Corbett*¹⁶ is not binding precedent in Australian law, and despite the contemporary legal departures from the biological criteria set up by Ormrod J, its influence continues even today. Although it is no longer cited favourably and its approach to sex identification is no longer followed, the mere fact that a case decided forty years ago in a different jurisdiction continues to even be discussed is testament to its importance in this area of law.¹⁷ Contemporaneous legal positions and arguments are still jurisprudentially framed as oppositional in relation to the restrictive biological model contained in *Corbett v Corbett*.¹⁸

Just as important as the jurisprudential effect of *Corbett v Corbett*¹⁹ was the practical legal effect of Ormrod J’s test with regards to transsexual subjects. When ‘all the external social indicia of a human relationship [fall] away, melted by the discovery of a genetic pattern, marked before birth, but demonstrable only by peering down a microscope’,²⁰ the ultimate result is that a transsexual person is powerless to display their psychological sex identity in social, psychological or (contemporaneous) biological terms that impact on legal decision-making. Tobin

⁹ Ibid. That is, a ‘man’ must have XY chromosomes, testes, and a penis and testicles, whereas a ‘woman’ must have XX chromosomes, ovaries, and a vulva, vagina and (possibly) uterus.

¹⁰ Ibid.

¹¹ Ibid 104.

¹² [1971] P 83.

¹³ [1983] QB 1053.

¹⁴ [1971] P 83.

¹⁵ *R v Tan* [1983] QB 1053, 1064.

¹⁶ [1971] P 83.

¹⁷ See, eg, *Secretary, Department of Social Security v SRA* (1993) 43 FCR 299; *AB & AH v Western Australia* (2011) 244 CLR 390.

¹⁸ [1971] P 83.

¹⁹ Ibid.

²⁰ Michael Kirby, ‘Foreword’ in H A Finlay and William Walters, *Sex Change: Medical and Legal Aspects of Sex Reassignment* (H A Finlay, 1988) [viii].

describes Ormrod J's strict born biological criteria as a 'doctrine of immutability', because a subject is legally fixed forever in the sex assigned to them by their biology at birth, and criticises its 'far-reaching and pernicious effects for transsexual people and their families'.²¹ It is self-evident that this legal approach negates the access of transsexuals to legal recognition of their psychological sex identity. Sex reassignment procedures, hormone therapy, socialisation and psychological identification may be meaningful life projects for transsexual subjects but, under this test, they have no impact on their legal treatment. Transsexual subjects may have significantly altered their bodies and their lives, but *Corbett v Corbett*²² denied them any legal recognition of these changes by trapping them within rigid, legal sex identifications pre-ordained at their birth.

B *Psychological and Anatomical Harmony*

Australian law broke with Ormrod J's strict biological criteria in *R v Harris and McGuinness*,²³ a case that adopted a slightly broader legal test that was reliant upon psychological self-identification in addition to looser biological criteria. This case was concerned with two transwomen charged with prostitution-based criminal offences. An element of the statutory charges each transwoman faced was that they were 'a male person' at the time of the relevant conduct. Importantly for the decision, whilst both Harris and McGuinness were born biologically male and identified and lived socially as women, Harris had undergone genital sex reassignment surgery whereas McGuinness had not. In a 2:1 decision, the majority refused to follow *Corbett v Corbett* and decided that Harris was no longer legally male.²⁴ Mathews J, in her leading judgment (with which Street CJ agreed), declared that: '[t]he time, then, has come when we must, for the purposes of the criminal law, give proper legal effect to successful reassignment surgery undertaken by transsexuals'.²⁵ In giving legal effect to such surgeries, Mathews J emphasised their impact on the sexual and procreative capacities of reassigned transsexuals — she noted that such sex reassignment surgery

²¹ Harper Jean Tobin, 'Against the Surgical Requirement for Change of Legal Sex' (2006–2007) 38(2) *Case Western Reserve Journal of International Law* 393, 404.

²² [1971] P 83.

²³ (1988) 17 NSWLR 158.

²⁴ However, Carruthers J approved of and followed *Corbett v Corbett* [1971] P 83 in his dissenting judgment. In response to evidence about Harris' and McGuinness' psychological identification as women, he concluded, at [158], that:

The law could never countenance a definition of male or female which depends on how a particular person views his or her own gender. The consequence of such an approach would be that a person could change sex from year to year despite the fact that the person's chromosomes were immutable ...

[T]he position is more complex in the case of transsexuals who have undergone reassignment surgery such as the appellant Lee Harris. In essence, the surgery in her case involved the removal of the penis and the creation of a cavity which was intended to act as a substitute vagina. However, such surgery combined with "her" hormone therapy and "her" psychological attitude to her gender cannot possibly, to my mind, override the congruence of the chromosomal, gonadal and genital factors which are all male.

²⁵ *R v Harris* (1988) 17 NSWLR 158, 181.

‘permanently deprives the [patient] of the capacity to procreate or to have normal heterosexual intercourse in her original sex’²⁶ — as well as their appearance — such surgery ‘generally produces a person who bears the external features of [the other sex]’.²⁷ The overall goal for such surgery, and the key factor that should be given effect by the law, is that ‘the body is brought into harmony with the psychological sex’.²⁸ Although McGuinness psychologically identified as being female, her failure to undergo sex reassignment surgery constituted a failure to establish this harmony, and her psychological sex identification was held to be an insufficient basis by itself for legal recognition to be granted to her female sex identity.²⁹

The next major Australian case to deal with sex identification law was *Secretary, Department of Social Security v SRA*.³⁰ This case was concerned with determining whether a transwoman fell under the ordinary meaning of the word ‘woman’ for the purposes of determining her entitlement to a pension based on being the ‘wife’ of an invalid pensioner. Sharpe identifies the legal shift away from strict biological criteria to a model based on ‘psychological and anatomical harmony’ as finding purchase in this case.³¹ Indeed, the Court in *Secretary, Department of Social Security v SRA*³² ultimately seems to approve of and apply the approach taken in *R v Harris and McGuinness*.³³ However, in the course of deciding that SRA was not entitled to such a pension, Lockhart J departed from the specifics of Mathews J’s wording from *R v Harris and McGuinness*³⁴ when he noted that:

The principal difficulty which I have in this case is ... the recognition of a pre-operative transsexual as being a member of the adopted sex for the purposes of the law ... [S]uch a person has not *harmonised her anatomical sex and her social sex*; they are not in conformity. She still has the genitals of a man.³⁵

²⁶ Ibid 180. The concern here that Mathews J displays for ‘normal’ sex as well as for ‘heterosexual intercourse’ is obviously problematic in that it devalues alternative sexualities and enshrines normative heterosexuality in a privileged position.

²⁷ Ibid.

²⁸ Ibid.

²⁹ Ibid 181. A plight which Mathews J had ‘the greatest sympathy for’, but she felt that she could not adopt psychological sex identification as the only criteria for legal recognition because it ‘would create enormous difficulties of proof’, it ‘would be vulnerable to abuse by people who were not true transsexuals at all’, and ‘it could lead to a trivialisation of the difficulties genuinely faced by people with gender identification disharmony’.

³⁰ (1993) 43 FCR 299.

³¹ Andrew Sharpe, ‘From Functionality to Aesthetics: The Architecture of Transgender Jurisprudence’ (2001) 8(1) *eLaw Journal: Murdoch University Electronic Journal of Law* [7] <<http://www.murdoch.edu.au/elaw/issues/v8n1/sharpe81.html>>.

³² (1993) 43 FCR 299.

³³ (1988) 17 NSWLR 158.

³⁴ Ibid.

³⁵ *Secretary, Department of Social Security v ‘SRA’* (1993) 43 FCR 299, 326 (emphasis added).

Although Lockhart J's judgment addressed psychological sex identification at numerous points, and appears to place great emphasis on it in addition to biological factors, this passage subtly but importantly alters the wording of Mathews J's concept of 'biological and psychological harmony' by replacing it with 'anatomical and social harmony'. Although nothing within the decision turns on the distinction between 'psychological' and 'social', the wording of Lockhart J's judgment prefigures what this article identifies as the rise of social factors as important legal considerations in sex identification law. This is made most clear in an additional comment he made as an introductory obiter dicta remark: '[s]ex is not merely a matter of chromosomes, although chromosomes are a very relevant consideration. Sex is also partly a psychological question (a question of self-perception) and partly a social question (how society perceives the individual)'.³⁶

Although not yet marking a full turn towards the multifactorial approach that was to be adopted in Australian law, this test of biological and psychological harmony greatly mitigated the harshness of Ormrod J's strict biological criteria. Transsexual subjects could now be legally recognised as a sex other than their born biological sex. The doctrine of immutability had been replaced, but whilst the law now allowed individuals to be reallocated across the male or female legal sex identification divide, it still utilised strict criteria to control such reallocations. Both *R v Harris and McGuinness*³⁷ and *Secretary, Department of Social Security v SRA*³⁸ required transsexuals to undergo sex reassignment surgery before granting legal recognition of their psychological sex identification. This surgical requirement sets up practical barriers for transsexual subjects who cannot afford such extensive procedures, who are medically unable to undergo major surgery (due to illness, age or likely medical complications), who do not want to run the physical risks involved, or who simply feel that such surgery is unnecessary for them to identify with their psychological sex identity.³⁹

C Multifactorial Approach

The turn of the century was accompanied by a significant shift in Australian law's treatment of the sex identification of transsexuals. The landmark decision in *Re Kevin: Validity of Marriage of Transsexual*⁴⁰ not only stands as authority for the explicit rejection of *Corbett v Corbett*⁴¹ in relation to Australian marriage

³⁶ Ibid 325.

³⁷ (1988) 17 NSWLR 158.

³⁸ (1993) 43 FCR 299.

³⁹ Indeed, 'many transsexual, transgender, and gender nonconforming individuals find comfort with their gender identity, role, and expression without surgery': The World Professional Association for Transgender Health, *Standards of Care for the Health of Transsexual, Transgender, and Gender Nonconforming People* (7th ed, 2012) 54.

⁴⁰ (2001) 165 FLR 404.

⁴¹ [1971] P 83.

law,⁴² but also stands for the adoption of a multifactorial approach to determining sex that encompasses a wide variety of considerations. In the course of finding that Kevin, a transman who had undergone partial sex reassignment surgery,⁴³ was ‘male’, Chisholm J concluded that:

To determine a person’s sex for the purpose of the law of marriage, all relevant matters need to be considered ... [R]elevant matters include, in my opinion, the person’s biological and physical characteristics at birth (including gonads, genitals and chromosomes); the person’s life experiences, including the sex in which he or she is brought up and the person’s attitude to it; the person’s self-perception as a man or woman; the extent to which the person has functioned in society as a man or a woman; any hormonal, surgical or other medical sex reassignment treatments the person has undergone, and the consequences of such treatment; and the person’s biological, psychological and physical characteristics at the time of the marriage, including (if they can be identified) any biological features of the person’s brain that are associated with a particular sex.⁴⁴

An appeal to the Full Federal Court in *Attorney-General (Commonwealth) v Kevin*⁴⁵ was dismissed. In relation to the argument on appeal that Chisholm J had erred in taking social factors into account when determining Kevin’s sex, the Full Court found that Chisholm J was ‘correct in paying attention to the evidence as to social and cultural factors’,⁴⁶ and that it was ‘clearly relevant to receive evidence as to how Kevin and [his wife] are perceived by the community in which they live’.⁴⁷ In a new twist on the law relating to sex identification, in addition to biological and psychological factors the law now accepted that ‘society’s perception of [a] person’s sex provides relevant evidence as to the ordinary, everyday meaning of the words “man” and “woman”’, and thus whether or not a transsexual fulfils the criteria to have their psychological sex identity legally recognised.⁴⁸

⁴² Indeed, Chisholm J could not be more explicit when he states in *Re Kevin: Validity of Marriage of Transsexual* (2001) 165 FLR 404, 474–5 [326], that:

I see no basis in legal principle or policy why Australian law should follow the decision in *Corbett*. To do so would, I think, create indefensible inconsistencies between Australian marriage law and other Australian laws. It would take the law in a direction that is generally contrary to developments in other countries. It would perpetuate a view that flies in the face of current medical understanding and practice. Most of all, it would impose indefensible suffering on people who have already had more than their share of difficulty, with no benefit to society.

⁴³ Whilst Kevin had removed his internal female genitalia, such as his uterus and ovaries, he had not had male genitalia surgically constructed.

⁴⁴ *Re Kevin: Validity of Marriage of Transsexual* (2001) 165 FLR 404, 475 [329].

⁴⁵ (2003) 172 FLR 300.

⁴⁶ *A-G (Cth) v Kevin* (2003) 172 FLR 300, 336.

⁴⁷ *Ibid* 329.

⁴⁸ *Ibid* 330.

Around the same time period, a slew of State and Territory legislation codified sex identification law relating to birth certificates and other legal documentation.⁴⁹ The requirements they set out vary from jurisdiction to jurisdiction, but can be divided into two broad categories: the sex reassignment surgical requirements in the Australian Capital Territory, Northern Territory, Queensland, Tasmania, Victoria and New South Wales, and the multifactorial requirements in Western Australia and South Australia. The requirements for the former group of jurisdictions are narrowly focused on whether or not a transsexual person has undergone a sex reassignment procedure.⁵⁰ These requirements seem to be a pared back version of the psychological and anatomical harmony test; they mostly jettison the psychological requirements but retain the focus on sex reassignment surgeries as being sufficient to grant legal recognition of a transsexual's psychological sex identity. The requirements for the latter group of jurisdictions are more numerous and more detailed, and refocus legal consideration on factors that had previously been overlooked.

The *Gender Reassignment Act 2000* (WA) was highly influenced by the *Sexual Reassignment Act 1988* (SA), to the extent that they share identical wording in certain sections because parts of the earlier Act were copied wholesale into the later Act. Both jurisdictions require that an applicant for a recognition certificate (a certificate that functions as 'conclusive evidence' that the applicant is the sex noted in the recognition certificate)⁵¹ has received counselling⁵² and that they have undergone a 'reassignment procedure'.⁵³ The important issue for this article, however, is the additional requirement that the governing body (a magistrate in South Australia, or the Gender Reassignment Board in Western Australia) must be satisfied that the applicant:

- believes that his or her true sex (SA) or gender (WA) is the sex (SA) or gender (WA) to which the person has been reassigned; and

⁴⁹ See, eg, the *Births, Deaths and Marriages Registration Act 1997* (ACT); *Births, Deaths and Marriages Registration Act 2005* (NT); *Births, Deaths and Marriages Registration Act 2003* (Qld); *Births, Deaths and Marriages Registration Act 1999* (Tas); *Births, Deaths and Marriages Registration Act 1996* (Vic); *Births, Deaths and Marriages Registration Act 1995* (NSW); *Gender Reassignment Act 2000* (WA). The early forerunner of this kind of legislation was the *Sexual Reassignment Act 1988* (SA).

⁵⁰ To use Queensland as an example, under s 23(4)(b)(i) of the *Births, Deaths and Marriages Registration Act 2003* (Qld) all that is needed for a successful application to change one's birth certificate sex identification is an application accompanied by statutory declarations from two doctors verifying that the person has undergone sexual reassignment surgery. 'Sexual reassignment surgery' is defined by sch 2 of the same Act to mean 'a surgical procedure involving the alteration of a person's reproductive organs'.

⁵¹ *Sexual Reassignment Act 1988* (SA) s 8; *Gender Reassignment Act 2000* (WA) s 16.

⁵² *Sexual Reassignment Act 1988* (SA) s 7(8)(b)(iii); *Gender Reassignment Act 2000* (WA) s 15(1)(b)(iii).

⁵³ *Sexual Reassignment Act 1988* (SA) s 4; *Gender Reassignment Act 2000* (WA) s 14(1).

- has adopted the lifestyle and has the sexual (SA) or gender (WA) characteristics of a person of the sex (SA) or gender (WA) to which the person has been reassigned.⁵⁴

The first limb of this test simply replicates the familiar legal requirement that an applicant's psychological sex identification must be in 'harmony' with the biological criteria required by the law, but the second limb of this test introduces social factors as new legal considerations. Consideration of 'lifestyle' clearly extends the relevant legal factors beyond the mere biology and psychology of a transsexual applicant, as does the requirement relating to sexual or gender characteristics because it is statutorily defined as meaning 'the physical characteristics by virtue of which a person is *identified* as male or female'.⁵⁵

In *AB & AH v Western Australia*⁵⁶ the High Court defined what, exactly, these new considerations comprised. This case concerned two transmen who had made unsuccessful applications to receive recognition certificates under the *Gender Reassignment Act 2000* (WA). In deciding whether or not previous adjudicating bodies (including the Gender Reassignment Board, the State Administrative Tribunal and the Western Australian Court of Appeal) had correctly applied the relevant statute law, the High Court decided the legal meaning of the 'lifestyle' and 'gender characteristics' statutory considerations. With regards to the 'lifestyle' requirement, the High Court noted that:

The word 'lifestyle' refers to the characteristic manner in which a person lives and reflects a collection of choices which that person makes. It has both a private and a public dimension. Many lifestyle choices made by a person are observable by other members of society, by reference to how that person lives and conducts himself or herself. The first enquiry ... may therefore also direct the attention of the [governing body] to a social perspective.⁵⁷

With regards to 'gender characteristics', the High Court decided that:

The question whether a person is identified as male or female, by reference to the person's physical characteristics, is ... largely one of social recognition. It is not intended to require an evaluation by the [governing body] of how much of a person's body remains male or female. Rather, the [governing body] is directed ... to the question of how other members of society would perceive the person, in their day-to-day lives.⁵⁸

⁵⁴ *Sexual Reassignment Act 1988* (SA) s 7(8)(b)(i)–(ii); *Gender Reassignment Act 2000* (WA) s 15(1)(b)(i)–(ii).

⁵⁵ *Sexual Reassignment Act 1988* (SA) s 3; *Gender Reassignment Act 2000* (WA) s 3 (emphasis added).

⁵⁶ (2011) 244 CLR 390.

⁵⁷ *AB & AH v Western Australia* (2011) 244 CLR 390, 403 [28].

⁵⁸ *Ibid* 405 [35].

Legal consideration of these social factors marks a distinct departure from the jurisprudential history discussed above. Various contemporary sources of law have embraced a multifactorial approach that seems to broadly consider transsexual subjects and their lives rather than reductively focusing solely on transsexual bodies. In this legal development Whittle identifies a shift from ‘essentialist’ legal tests, where the law employs ‘tests that look to one essential feature’ and assigns a legal sex identity accordingly, to a ‘cluster’ approach, where the law ‘looks to a group of similar features that suggest’ the allocation of a legal sex identity.⁵⁹ With the introduction of additional factors for consideration, each individual factor becomes relatively less important to the entire decision-making process. Under this new ‘cluster’ approach, biology remains a factor to be taken into account, but is stripped of its power to solely determine the outcome of legal decision-making.

III PROBLEMS WITH THE USE OF SOCIAL FACTORS

A general trend that cuts across many of the changes made within sex identification law in the last forty years is the opening up of legal consideration from the narrow confines of born biological criteria to a variety of other considerations, including contemporaneous biology (caused by hormonal and surgical treatments), psychological identification and social factors. By no means, however, is this multifactorial approach the only, or even the dominant, legal test in Australia. The multifactorial approach has only been wholeheartedly adopted within marriage law and within Western Australian and South Australian law regarding sex identifying documentation.⁶⁰ Other areas of sex identification law have different foci. The prerequisites required to change the sex noted in a birth certificate in other jurisdictions retain a narrow focus on biological criteria: typically just requiring evidence of undergoing sex reassignment surgery.⁶¹ Recent changes to the federally-administered passport sex identification policy mean that passport criteria now rely on the professional judgment of medical practitioners. All that a transsexual needs to change the sex identity noted on their passport is a letter from a medical practitioner (registered

⁵⁹ Stephen Whittle, *Respect and Equality: Transsexual and Transgender Rights* (Cavendish Publishing Limited, 2002) 10–11.

⁶⁰ And perhaps also in the law regarding social security. See, eg, *Scafe v Secretary, Department of Families, Housing, Community Services and Indigenous Affairs* (2008) 100 ALD 131, [21] (*‘Scafe’*) wherein the Tribunal decided that:

There is much to be said for the view that, in reaching a conclusion as to the gender of an individual, consideration should be given to and a determination made in light of all the characteristics of that person, including behavioural and psychological matters and social circumstances. The individual should be evaluated as a complete human being, taking into account their full range of behaviour, physiology, psychology and any other relevant features and characteristics.

⁶¹ See, eg, the *Births, Deaths and Marriages Registration Act 1997* (ACT); *Births, Deaths and Marriages Registration Act 2005* (NT); *Births, Deaths and Marriages Registration Act 2003* (Qld); *Births, Deaths and Marriages Registration Act 1999* (Tas); *Births, Deaths and Marriages Registration Act 1996* (Vic); *Births, Deaths and Marriages Registration Act 1995* (NSW).

with the Medical Board of Australia) supporting the fact that they have ‘had, or [are] receiving, appropriate clinical treatment for gender transition’.⁶² Obviously, the law in these areas has no regard for social factors, and is far more responsive to issues of medical treatment and practice.

My contention, however, is that important parts of sex identification law in Australia are trending towards the multifactorial approach,⁶³ and, as a result, it is important to subject this trend to close analytical scrutiny. The multifactorial approach, as the names suggests, takes into account a number of factors about a transsexual person’s life, mind and body. My intention in this article, however, is just to scrutinise the use of social factors within the multifactorial approach. The consideration of biological and psychological factors are also key parts of the multifactorial approach but will not be addressed in detail here. Such factors have already received high levels of judicial and academic attention by virtue of having been relevant legal considerations for a much longer time.

It is tempting to locate a linear pattern of liberalising legal progression in the shift from biological essentialism to the multifactorial approach. The immutability of Ormrod J’s narrow biological criteria, and the insurmountable barriers it set up to the legal recognition of a transsexual’s psychological sex identity, has given way to a broad approach that considers a variety of factors and that is, on the whole, far less restrictive. However, my concern, and the major contention of this article, is that the introduction of social factors as relevant legal considerations could generate new barriers for transsexuals seeking legal recognition of their psychological sex identity that could be applied in inequitable ways. These factors, depending on how they are used by future courts and legal decision-makers, have the capacity to harden into a restrictive vector of normative legal control by becoming not simply an additional set of legal factors to consider but an additional set of legal requirements that transsexuals must overcome. This article constitutes a warning about possible future progress in this legal area; it aims to reveal the hidden dangers that litter this area of legal development by modelling some of the problematic ways in which the law could end up utilising these social factors. This argument is necessarily somewhat speculative, but it is not, however, purely hypothetical. Particular reference will be made to *Re Kevin: Validity of Marriage of Transsexual*⁶⁴ to demonstrate that the way these social factors have already been employed indicates that they may trade in restrictive and highly normative understandings about sex that are problematic for transsexuals seeking legal recognition of their psychological sex identities.

⁶² Australian Passport Office, *Sex and Gender Diverse Passport Applicants* (2011) Australian Government Department of Foreign Affairs and Trade <<https://www.passports.gov.au/Web/SexGenderApplicants.aspx>>.

⁶³ This is especially evident in the way the High Court read social factors into the statutory requirements of the *Gender Reassignment Act 2000* (WA) in *AB & AH v Western Australia* (2011) 244 CLR 390.

⁶⁴ (2001) 165 FLR 404.

The specific problems to be discussed in this part are threefold. Firstly, how the consideration of social factors undermines the importance of transsexuals' own psychological sex identifications and subjective experiences. Secondly, how these factors bring extensive and intrusive legal interrogatory pressure to bear on transsexuals' lives. Finally, how the implementation of these factors into legal decision-making could require that transsexuals conform to narrow, stereotypical models of sex.

A Devaluation of Transsexual Experience or Identification

Transsexualism is a profoundly personal experience, in the sense that it is defined by a person's strong and ongoing psychological identification with the sex other than their born biological sex,⁶⁵ as well as in the sense that surgical and medical treatment of transsexualism is undertaken to mitigate the subjective distress and anxiety that their incongruent sex identification can cause to some transsexuals.⁶⁶ The law around sex identification, however, relates to thoroughly public categories. Cruz argues that legal sex identities 'should be understood as relationships among classes of people',⁶⁷ and even Ormrod J recognised that '[t]he fundamental purpose of [this area of] law is the regulation of the relations between persons, and between persons and the state or community'.⁶⁸ The multifactorial approach radically refocuses the law's basis for assigning individuals to these public legal sex identities by placing less emphasis on the private, personal experience of the transsexual seeking legal recognition of their psychological sex identity, and by placing more emphasis on society's experience of that person's public expressions of their transsexualism.

If the law employs 'male' and 'female' as public categories that govern social relationships, what role should a transsexual person's private experience and self-identification have in deciding their legal sex? The legal shift away from biological essentialism in favour of the multifactorial approach also marks a shift away from the transsexual as the sole site of legal consideration and onto extraneous issues such as how that person's sex identity is perceived by others. Whatever else one has to say about *Corbett v Corbett*,⁶⁹ it should at least be noted that Ormrod J's sole concern was with the biological experiences and reality of April Ashley Corbett. In sharp contrast, Chisholm J in *Re Kevin: Validity of Marriage of Transsexual*⁷⁰ almost

⁶⁵ Isis Dunderdale, 'The Human Rights of Transsexuals' (1992) 17 *Alternative Law Journal* 23, 23.

⁶⁶ Tobin, above n 21, 399. Tobin comments that '[f]or many trans people, [sex reassignment surgery] is essential to achieving peace of mind and a successful life in their authentic gender; the alternative is constant anxiety, social maladjustment, depression, and the danger of suicide'.

⁶⁷ David B Cruz, 'Sexual Judgments: Full Faith and Credit and the Relational Character of Legal Sex' (2011) 46 *Harvard Civil Rights–Civil Liberties Law Review* 51, 54.

⁶⁸ *Corbett v Corbett* [1971] P 83, 105. This passage is even cited favourably, and reproduced in its entirety, in *R v Harris* (1988) 17 NSWLR 158, 180.

⁶⁹ [1971] P 83.

⁷⁰ (2001) 165 FLR 404.

seems to sideline Kevin by considering and extracting copious amounts of evidence not about Kevin's own sex identification but about other witnesses' experience of Kevin and what sex identity they allocated to him. Chisholm J considered evidence from 39 witnesses about how Kevin is socially 'referred to and treated', from individuals or organisations such as his fellow employees, his step-family, his childhood friends, and various clubs.⁷¹ Chisholm J treats this evidence as important because:

It shows [Kevin] as perceived by those involved with him in his family, at work, and in the community ... It shows him living a life that those around him perceive as a man's life. They see him and think of him as a man, doing what men do. They do not see him as a woman pretending to be a man. They do not pretend that he is a man, while believing he is not.⁷²

Chisholm J's rhetoric here seems laudable, his intent is to consider Kevin in the fullness and richness of his life and its context. However, legal consideration of these social factors could potentially be very problematic if applied to cases that are unlike Kevin's.

Drawing legal focus away from the transsexual person and placing importance on the opinions and reactions of other people relatively devalues the importance of that transsexual person's subjective experience and exposes the success of their legal claim to the whims of those around them. Consideration of social factors raises the concern that a transsexual person's 'self-perception [can be] potentially trumped by the perceptions of others in the legal determination of sex itself'.⁷³ Imagine a transman not unlike Kevin who lived in a small, conservatively-minded country town where the majority of the population was (for some reason or another) ideologically-inclined against transsexualism, believing perhaps that it was simply an immoral 'lifestyle' choice, a debilitating psychopathology or a 'sin' against a particular deity. If the community around this transman perceived him as an immoral, sick or sinful *woman*, rather than as a man, then the application of social recognition as a relevant legal consideration would damage that transsexual's claim for legal recognition of his male sex identity. That a transsexual person's access to legal rights and benefits could be held hostage by the potentially bigoted attitudes of those around them is a strong argument for placing little, if any, legal weight on social recognition of their sex identity. Placing too much weight on the consideration of the social recognition of a transsexual's sex identity 'might be viewed as creating a legal space for the [possible] reproduction of oppression of a marginalised group'.⁷⁴

A similar argument can be made with another thought-experiment. Imagine a transman who, despite his best efforts to the contrary during transitioning, still

⁷¹ *Re Kevin: Validity of Marriage of Transsexual* (2001) 165 FLR 404, 412 [37].

⁷² *Ibid* 417 [68].

⁷³ Andrew Sharpe, 'Thinking Critically in Moments of Transgender Law Reform: *Re Kevin and Jennifer v Attorney-General for the Commonwealth*' (2002) 11 *Griffith Law Review* 309, 324.

⁷⁴ *Ibid* 325.

physically appears to be female to those around him. Despite doing everything in his power to alter his behaviour and appearance, the people in his life still see him 'as a woman pretending to be a man'.⁷⁵ This thought-experiment reveals two problematic issues with the legal consideration of social factors. Firstly, consideration of social factors takes determinative power away from the transsexual whose life and rights are most profoundly affected by the legal question being decided and places that power in the hands of other people who, whilst possibly having a theoretical interest in the upholding of male or female as social categories generally, have no direct interest at stake in the determination of that transsexual's legal sex identity. Despite the liberalising gloss of the multifactorial approach considering transsexual subjects and lives holistically, these social factors could potentially be employed in profoundly disempowering ways. Secondly, rather than providing a transsexual seeking legal recognition of their psychological sex identity with another avenue to demonstrate the legal validity of their sex identity, the legal consideration of social factors could instead be simply setting up an additional requirement based on 'passing'. Being widely socially recognised as another sex is difficult; successful 'passing' requires onerous and ubiquitous displays of sex-appropriate biology, affect and clothing (all of which are inflected by normative understandings about sex — as discussed below). Kevin's ability to effectively and consistently display masculinity to his friends, family and community cannot speak for the (in)ability of other transsexuals to perform the complex task of unambiguously blending into the social fabric as a member of their psychological sex identity.

Taking the focus away from a transsexual person and placing it on society also devalues that person's subjective experience of their sex identity. Such a factor should be of fundamental importance not only in determining whether their psychological sex identity should be legally recognised but also in weighing up other factors under legal consideration. An underlying concern within Martin CJ's judgment in *Western Australia v AH*,⁷⁶ before the case reached the High Court, was the fact that the two transmen seeking recognition certificates retained their internal female reproductive organs, including their uterus and ovaries which, if they ceased hormone treatment, could possibly regain functionality. Whilst Martin CJ made it clear that 'functionality or fertility is not the sole determinant of any application for a recognition certificate', he also noted that their possible capacity to still 'bear children' in the future was 'not irrelevant to the process of evaluation which falls to be undertaken'.⁷⁷ This anxiety over legally allowing for the creation of a 'pregnant man' is underpinned by a normative conception that gestation and parenting is, always and essentially, 'female' or 'feminine' work. Subjective, non-legal accounts from transmen who have become pregnant, or who have born children either during or soon after transitioning, dispute this normative association. It was reported by More that such transmen make efforts to maintain their male sex identities, and communicate their masculinity to others, including medical personnel, despite their

⁷⁵ *Re Kevin: Validity of Marriage of Transsexual* (2001) 165 FLR 404, 417 [68].

⁷⁶ (2010) 41 WAR 431.

⁷⁷ *Ibid* 458 [112].

pregnancy.⁷⁸ They also tend to characterise their feelings towards their unborn children as paternal rather than maternal.⁷⁹ A transman named Ben, when talking about the possibility of becoming pregnant again, told More that he ‘wouldn’t be less of a man because of it’.⁸⁰ The point is that a transsexual subject’s own experience of their sex identification is an important source of meaning that should be heavily legally weighted. Relying on normative cultural accounts of things like parenthood fails to appreciate the subjective accounts of transsexuals, whose own meanings about issues related to sex identification may contest and even directly contradict these normative accounts. While Martin CJ saw the transmen’s capacity for pregnancy and childbirth as fundamentally inconsistent with their masculinity, it could very well be that these capacities could be subjectively considered by a transman to support, rather than undercut, their masculinity, because it gives them the subsequent opportunity to engage with their biological child as a ‘father’. The law should be sensitive and respectful towards the subjective experiences and accounts of the transsexuals whose lives are regulated by this area of law, rather than devaluing them in the course of considering wider social factors under the multifactorial approach.

B *Intrusive Legal Interrogatory Pressure*

The biological criteria employed by Ormrod J in *Corbett v Corbett*⁸¹ narrowly engaged with transsexual subjects; legal focus was turned solely and specifically towards a small set of biological markers during a small temporal frame (ie, birth and soon after birth). The legal turn towards the multifactorial approach drastically increases the scope of this focus, potentially proliferating the sites at which transsexual subjects are interrogated by the law and increasing the level of legal scrutiny over their lives. Transsexual subjects may need to work harder than ever to achieve legal recognition of their psychological sex identity, in order to satisfy an increasingly wide-ranging legal analysis that draws on their entire life-histories and social connections as evidentiary sources.

The jurisprudential seeds for this broad focus on the life history of transsexuals were laid as early as *Corbett v Corbett*.⁸² There is more than a hint of irony in the declaration from Ormrod J that ‘the relevant facts must now be stated *as concisely as possible*’,⁸³ given that he spent the next seven pages of his judgment tracking April Ashley Corbett’s life from her birth in Liverpool, through her years in the Merchant Navy, to her stint as a ‘female impersonator’ at a cabaret in Paris, and finally to the details of her romance with her husband (including full extracts of letters to each

⁷⁸ Sam Dylan More, ‘The Pregnant Man — An Oxymoron?’ (1998) 7 *Journal of Gender Studies* 319.

⁷⁹ *Ibid* 322.

⁸⁰ *Ibid* 324.

⁸¹ [1971] P 83.

⁸² *Ibid*.

⁸³ *Ibid* 89 (emphasis added).

other).⁸⁴ The irony only deepens with the realisation that Ormrod J considers all this material to be legally irrelevant, because it is not based solely around genitals, gonads and chromosomes. This kind of unnecessary legal recounting of the fine details of transsexuals' lives recurs in other cases. Although proposing a test based strictly on harmony between biological and psychological factors, Mathews J in *R v Harris*⁸⁵ segues part of her judgment into a consideration of Harris' evidence that her mother had wanted her to be born a girl instead and 'used to dress her, as a child, in her sisters' clothes', and that Harris had spent her work life in stereotypically feminine jobs, including being 'a hairdresser and a seamstress'.⁸⁶ These narrative details, although legally unnecessary and technically irrelevant to the ultimate decisions in these cases, provide authorising accounts about the authenticity of transsexual lives: they purport to show unbroken patterns of historical identification and behaviour that support transsexuals' contemporaneous legal claims about their psychological sex identity. They operate as background notes to these legal judgments that, whilst perhaps not determinatively influencing the decisions, position the sex identity claims of the transsexuals before the courts as sympathetic and authentic.

The discussion of transsexual life-histories in sex identification law becomes problematic when, under the multifactorial approach, they shift from the background into the foreground of legal consideration, possibly providing restrictive models of how 'proper' transsexual lives need to progress in order for sex identity claims to be successful. The consideration of Kevin's life history in *Re Kevin: Validity of Marriage of Transsexual*⁸⁷ provides a good example. Chisholm J works through the evidence about Kevin's childhood, adolescence and life in extreme detail, making the following observations:

[F]or as long as he could remember, Kevin has perceived himself to be male. When he was a very young child his mother tried to persuade him that he was a girl and that he should behave as a girl ... [but] he continued to believe he was a boy. He wore boys' clothes whenever he could. He refused to play with girls' toys.⁸⁸

Kevin was the oldest of four children: he had three sisters. He saw his relationship with them as being that of an older brother. He would physically defend them, at school and elsewhere, after his father had left the family home. He did some of the physical tasks his father had done, such as mowing the lawns and doing household repairs.⁸⁹

Some family photographs are striking: at age three, with pistols; at age eight, with a soccer ball and trophy. Most remarkable is a photograph of Kevin aged about 15

⁸⁴ Ibid 89–96.

⁸⁵ (1988) 17 NSWLR 158.

⁸⁶ Ibid 161–2.

⁸⁷ (2001) 165 FLR 404.

⁸⁸ Ibid 410 [24].

⁸⁹ Ibid 410 [25].

or 16, with his sisters. They are wearing pastel coloured dresses and sandals. He is wearing dark trousers and shoes, and what looks like a boy's shirt. To my eye, despite the shoulder length hair, he looks as much like a boy as a girl.⁹⁰

In late 1994 he commenced work with his present employer. Throughout his employment there he generally presented as a male, wearing trousers and shirts to work.⁹¹

As part of his legal consideration of 'all relevant matters' in deciding Kevin's legal sex identity,⁹² Chisholm J delves into and considers all the obscure minutiae of Kevin's life: what chores he performed at home, what toys he played with as a child, what kind of shirts he and his sisters wore during his adolescence, what kind of pants he wore to his workplace seven years ago, etc. This kind of legal analysis *ad nauseum* extends interrogatory pressure into wide-ranging aspects of a transsexual subject's person and life. No longer is it just their bodies that are under intense legal scrutiny, so too are their childhoods, their adolescences, their sartorial choices and their workplaces. The notion of 'legal consideration of social factors' could merely be a neutral epithet placeholder for the law's self-grant of an exceedingly intrusive level of oversight (and control) over transsexual lives.

This vast increase in the scope and intensity of the legal interrogation of transsexual lives that the multifactorial approach brings to sex identification law is problematic in and of itself. It is important, however, to spend some further time unpacking the issues that this raises.

Firstly, the proliferation of social factors as relevant legal considerations correlatively proliferates the various sites through which transsexual subjects must demonstrate their masculinity or femininity to the law. Their life histories, social recognition and entire past catalogue of gendered displays take on evidentiary importance. As evidence, they can be used either to support or undermine their claims for legal recognition of their psychological sex identity. This becomes a concern for those transsexual subjects who, unlike Kevin, may have a chequered past filled with late-blooming or inconsistent sex-appropriate behaviour. Where does this leave the transman who played with 'girls' toys when he was three years old? What about the transwoman who looked like a 'boy' in her old family photographs? What about the transman who only started dressing in 'male' clothes later in life? The use of these social factors as legal considerations has the possibility to leverage significant and lifelong pressure on transsexuals to perform sex-appropriate behaviour, and display sex-appropriate appearance, during every single moment of their entire lifetime. Furthermore, there is the capacity for evidence of sex-inappropriate behaviour or appearance to be dredged from the most remote recesses of a transsexual's life-history and brought forward in opposition to their claim for legal recognition of their psychological sex identification.

⁹⁰ Ibid.

⁹¹ Ibid 411 [27].

⁹² Ibid 475 [329].

The concern is that the law's broad preoccupation with a transsexual's life-history will coalesce into a legal investigation into whether particular transsexual subjects conform with an ideal transsexual life-narrative, the content of which is informed by the experiences of those transsexuals who fit the 'classic transsexual pattern' and not those whose current psychological sex identity has previously surfaced (or been expressed) either ambiguously or sporadically throughout their life. Not all transsexuals will be able to fulfil this. As Lawrence recognises with regards to transwomen, not all individuals follow the "'classic" transsexual pattern' of being 'extremely feminine as children [and] extremely feminine as adults' and seeking sex reassignment surgery and hormone treatment early in life.⁹³ Others may only seek medical treatment 'in their 30s, 40s, 50s, or even later, after having lived outwardly successful lives as men', and their life-histories may show that they 'were not especially feminine as children ... [or] as adults, either'.⁹⁴ There is the potential here for the law to differentially privilege certain models of transsexuality over and above others and thus to function restrictively. Such a legal move would run counter to the prevailing trends in the medical treatment of transsexualism. The current edition of The World Professional Association for Transgender Health's *Standards of Care for the Health of Transsexual, Transgender, and Gender Nonconforming People* reports that 'many adolescents and adults presenting with gender dysphoria do not report a history of childhood gender conforming behavior',⁹⁵ and promotes an inclusive understanding of gender identity that recognises it as a unique experience not necessarily bound to social norms.⁹⁶

Secondly, the construction of numerous, onerous legal hurdles for transsexuals to overcome in order to achieve legal recognition of their psychological sex identity is undercut by the blatant legal double-standard it imposes; sex only becomes something that needs to be legally achieved when someone is transsexual and not when they are cisgendered.⁹⁷ Cruz's observation that '[m]ost people do not experience their sex or gender being legally called into question or contested',⁹⁸ reveals a fundamental hypocrisy at the core of sex identification law. The various tests that the law requires transsexual subjects to fulfil before being legally recognised as their psychological sex identity do not have to be fulfilled by cisgendered subjects in order for them to be legally recognised as the sex with which they psychologically identify. Under the law, recognition of psychological sex identity is something that must be achieved, with reference to various considerations and criteria, but it is only something that transsexuals must achieve. This same point can be argued in relation to any legal considerations relating to sex identification;

⁹³ Anne A Lawrence, 'Autogynephilia: A Paraphilic Model of Gender Identity Disorder' (2004) 8(1) *Journal of Gay & Lesbian Psychotherapy* 69, 70.

⁹⁴ *Ibid.*

⁹⁵ The World Professional Association for Transgender Health, above n 39, 12.

⁹⁶ *Ibid.* 9.

⁹⁷ The word 'cisgendered' refers to people who are born with a psychological sex identification that is congruent with their physiological sex identity markers.

⁹⁸ Cruz, above n 67, 66.

it is equally valid as an argument against the biological essentialism of Ormrod J as it is against the multifactorial approach adopted by Chisholm J. The adoption of the multifactorial approach does, however, exacerbate the situation because it is particularly egregious for the law to proliferate the considerations that transsexuals must fulfil when the cisgendered are not required to fulfil any. To illustrate this point take, for example, biological criteria. Under the psychological and anatomical harmony test in *R v Harris and McGuinness*,⁹⁹ a transwoman is required to demonstrate that she has female genitalia before being legally recognised as female. In contrast, a cisgendered woman who ‘lost these [biological] features by accident, or had them removed for medical reasons’ retains her legal identification as female and any legal rights gained thereby.¹⁰⁰ Under the multifactorial approach, this inequality is extended to a wider variety of social factors as well. The cisgendered woman who plays with ‘boys’ toys, who dresses in ‘boys’ clothes and who enacts a stereotypically masculine social role (whilst maintaining a female psychological sex identification) does not run the risk of being denied legal recognition of her female sex identity and her resultant legal status and rights. Sex identification law constructs recognition of psychological sex identity as something that transsexuals must strive for and achieve, yet the legal recognition of a cisgendered person’s psychological sex identity is something that does not need to be achieved and that cannot even be ‘failed’ out of by sex-inappropriate displays. The broader variety of ways in which the multifactorial approach requires transsexuals to struggle to achieve legal recognition of their psychological sex identity extends this inequality by broadening the gap between the growing scope of the legal interrogation of transsexual lives and the accommodating approach the law takes to the legal recognition of cisgender sex identity.

C Conformity with Sex Stereotypes

The concerns outlined within the last two sections of this article relate to matters of legal content: that is, consideration of social factors seems to add additional, new restrictive criteria to sex identification law. In this section, I want to add to these concerns by highlighting a further problem about legal methodology: that is, consideration of social factors also seems to set up further restrictions through the processes involved in applying and testing these factors. The problem here is that the practical process of judicially considering these social factors could devolve into mere capitulation with (and reification of) normative understandings of male or female that trade in restrictive models of sex and gender. In considering the life-history of a transsexual and the social recognition of their sex identity, legal logic and decision-making becomes susceptible to being drawn into the application of ‘outdated concepts of ideal “men” and “women” and of normative masculinity, femininity, and sexuality’.¹⁰¹

⁹⁹ (1988) 17 NSWLR 158.

¹⁰⁰ Tobin, above n 21, 425.

¹⁰¹ Tey Meadow, “‘A Rose is a Rose’: On Producing Legal Gender Classifications” (2010) 24 *Gender & Society* 814, 815.

This broad point is best highlighted by the Western Australian and South Australian statutory requirement that before a transsexual person can receive a recognition certificate they must satisfy the decision-maker that they have ‘adopted the lifestyle ... of a person of the sex (SA) or gender (WA) to which the person has been re-assigned’.¹⁰² In addition to constructing an additional legal criterion for transsexuals to fulfil, that is, that they have to prove that they have adopted a particular lifestyle, the methodology for identifying and deploying a coherent account of what exactly constitutes a male or female ‘lifestyle’ is also highly problematic. This legal criterion cannot practically be employed without recourse to reductive, normative and ultimately stultifying conceptions about what constitutes masculinity and femininity, or else it would have no substantive content as a test. The requirements of this test have yet to be given any legal specificity in case law, but typical understandings of what would constitute a sex-appropriate ‘lifestyle’ would associate it with normative understandings about things such as a person’s name, their sartorial choices, their (non-)use of makeup or jewellery, their hobbies, their (non-)involvement in domestic labour, their (non-) involvement in sport, their type of job, their affect, their sexual expression, etc. How is a court to weight the fact that a transman wears skirts occasionally? Or that a transwoman is heavily involved in playing rugby? Legal decision-makers could only possibly weight these factors by internalising broader cultural conceptions of sex-appropriate behaviour, such as those that conceive of skirt-wearing as a ‘feminine’ activity and playing contact sport as ‘masculine’.

The practical implementation of this legal test has the potential to import highly restrictive understandings of normative behaviour into the law. By considering social factors, the imposition of legal fantasies that ‘rely not only on concepts of what men and women are but also on notions of what they are meant to do’¹⁰³ could force transsexuals to choose between conforming with normative models of sex or gender or failing to be granted legal recognition of their psychological sex identity. Transsexuals who display ‘incorrect’, ambiguous or simply unorthodox sexed-behaviour risk legal delegitimation of their psychological sex identity.

*Re Kevin: Validity of Marriage of Transsexual*¹⁰⁴ provides an example of a case in which these theoretical concerns have been translated into legal reality. In an illuminating close analysis of the culturally specific gender dynamics at play in the decision, Aizura argues that Kevin’s success at enacting a normative Australian stereotype of masculine sex identity played ‘an important role in presenting him as a sympathetic public subject, of value to the community’.¹⁰⁵ Indeed, Chisholm J takes pains to collect and collate evidence for what appears to be the singular purpose of

¹⁰² *Sexual Reassignment Act 1988* (SA) s 7(8)(b)(ii); *Gender Reassignment Act 2000* (WA) s 15(1)(b)(ii).

¹⁰³ Tey, above n 101, 830.

¹⁰⁴ (2001) 165 FLR 404.

¹⁰⁵ Aren Z Aizura, ‘Of Borders and Homes: The Imaginary Community of (Trans)Sexual Citizenship’ (2006) 7 *Inter-Asia Cultural Studies* 289, 299.

portraying Kevin as an authentic ‘Aussie bloke’. His judgment extracts evidence showing that Kevin:

- was part of a nuclear family, having both a wife and children;¹⁰⁶
- played numerous sports such as rugby, soccer and cricket;¹⁰⁷
- went on fishing trips;¹⁰⁸
- moved and danced like a man;¹⁰⁹
- had a ‘bachelor-like’, Spartan home;¹¹⁰ and,
- was very practical and handy around the house, as he was skilled at both landscaping and household repair work.¹¹¹

Kevin’s ability to enact this ‘Aussie bloke’ identity is strategically legally valuable; by drawing on the weight of cultural narratives around (Australian) masculinity Kevin submerges the potential reading of his sex identity as that of a ‘gender freak’¹¹² by tapping into a ‘mythical space of normality’,¹¹³ both in terms of general Australian-ness and in terms of specific sex-appropriate behaviour as inflected by cultural custom.

Whilst the legal consideration of social factors may not have been problematic for Kevin (indeed, a strong argument could be made that consideration of such factors bolstered the strength of his claim for legal recognition of his psychological sex identity), it would be highly problematic for transsexuals generally if Chisholm J’s line of reasoning was to be extrapolated to other cases. The kind of masculinity embodied by Kevin is not only dated, it is also highly restrictive. It fixes certain types of behaviours, certain clothing, certain choices and certain ways of living in the world as fundamentally and essentially sexed, and this article’s concern is that these social practices will harden into legal requirements or guidelines that trap transsexuals into narrowly confined modes of authorised living. It is easy to envisage situations in which the realities of transsexual lives could contravene stereotypical cultural ideals of sex-appropriate behaviour. Take, for example, a transwoman who identifies as a lesbian. Dominant cultural understanding about femininity that link being female to androphilia, thus creating a nexus between

¹⁰⁶ *Re Kevin: Validity of Marriage of Transsexual* (2001) 165 FLR 404, 412 [38].

¹⁰⁷ *Ibid* 415 [55].

¹⁰⁸ *Ibid*.

¹⁰⁹ *Ibid* 415–16 [56]–[59].

¹¹⁰ *Ibid* 416 [59].

¹¹¹ *Ibid* 416 [59]–[60].

¹¹² Aizura, above n 105, 299.

¹¹³ *Ibid* 301.

sex and heteronormativity, are contravened here. The lesbian transwoman may fail to display (hetero)normative female behaviour, raising legal questions about how exactly this should be weighted in determining whether to grant legal recognition of her sex identification.¹¹⁴

This issue should be regarded as being symptomatic of law's broader essentialising of sex. The mere fact that there are any legal requirements that must be fulfilled before granting a person's wish to alter their legal sex identity demonstrates the law's complicity in constructing and maintaining sex identity categories, and the law's use of both biological and psychological legal requirements demonstrates its commitment to conceptualising fundamental differences between 'male' and 'female' bodies and minds. Legal consideration of social factors threatens to take this one step further: it runs the risk of legal decision-making committing the law to essentialist understandings about fundamental differences between 'male' and 'female' social identities, lives and lifestyles as well. Normative understandings about sexed embodiment have historically limited the access of transsexuals to legal recognition of their psychological sex identity and have required them to drastically change their bodies (such as through surgery). In a similar way, normative understandings about the sex-appropriateness of social identities and life-histories could limit the access of transsexuals to legal recognition and require them to drastically alter their lifestyles and lives.

IV CONCLUSION

This article is a caution, a warning sign indicating that the jurisprudence of sex identification law has approached difficult terrain that requires careful navigation. Part II detailed the shift from the biological essentialism of *Corbett v Corbett*¹¹⁵ towards the multifactorial approach, as endorsed in marriage law and in Western Australian and South Australian statute law. This expansion of the relevant factors for legal consideration has increased the access of transsexuals to legal recognition of their psychological sex identity by jettisoning the insurmountable biological criteria, and it also prima facie seems to allow the law to holistically account for complete and contextualised transsexual subjects. However, this legal shift also raises the concern that, upon closer analysis, the multifactorial approach may set up new legal barriers that could be applied inequitably to deny such recognition. Part III crystallised this concern into specific arguments about how the legal consideration of social factors

¹¹⁴ This very situation was considered in *Scafe* (2008) 100 ALD 131, where the Tribunal was tasked with deciding whether a lesbian transwoman who had undergone long-term hormonal therapy, but who had not undergone genital sex reassignment surgery, was female at law. Unfortunately, the Tribunal chose not to address the issue of the legal weight to be given to social factors, as it puzzlingly chose to follow the psychological and anatomical harmony approach from *Secretary, Department of Social Security v 'SRA'* (1993) 43 FCR 299 even though it seemed to endorse the multifactorial approach and explicitly stated that it thought there were 'grounds for distinguishing the circumstances and reasoning' of this case from 'SRA': *Scafe* (2008) 100 ALD 131, [29].

¹¹⁵ [1971] P 83.

about a transsexual's life could operate in disempowering and disenfranchising ways. Whilst these arguments were predominantly speculative, that is they identified *potentially* problematic areas within the jurisprudence, they were also grounded in the particular problems and approaches that have arisen from statute and case law that employs the multifactorial approach, with specific reference being made to *Re Kevin: Validity of Marriage of Transsexual*.¹¹⁶ Through its engagement with these problems, this article identifies a number of pitfalls inherent in the multifactorial approach, and provides a cautionary note about the need to proceed carefully and deliberately in both the development and the application of this area of law.

Sex identification law regulates a matter of intense concern for a small and highly marginalised group of people. As such, its development and application should occur on a principled basis, and should proceed with a high degree of clarity and coherence. This area of law has changed rapidly and extensively across the last forty years, and the multifactorial approach is one key contemporaneous trend in legal thinking. However, it is difficult to determine whether or not this latest development constitutes a positive step for transsexuals seeking legal recognition of their psychological sex identities. What is needed in this area of law now is not further development but further clarification. It is imperative to resolve the problems associated with the multifactorial approach so that it becomes clear whether these legal changes increase or decrease legal access, construct or dismantle legal restrictions, and ultimately progress or simply alter the law.

¹¹⁶ (2001) 165 FLR 404.

Michael Legg and Dean Jordan***

THE AUSTRALIAN BUSINESS JUDGMENT RULE AFTER *ASIC v RICH*: BALANCING DIRECTOR AUTHORITY AND ACCOUNTABILITY

ABSTRACT

The decision in *Australian Securities and Investments Commission v Rich* (2009) 236 FLR 1 (*ASIC v Rich*) has resurrected the business judgment rule in Australian corporate law, so that the rule is capable of providing a defence in some cases that would otherwise amount to a breach of a director's duty. This article reviews the controversial elements of the business judgment rule that were determined in *ASIC v Rich*, namely: (a) which party should bear the onus of proof in raising the defence; (b) what is required for a director to be informed about the subject matter of the judgment; and (c) is the standard that the director's judgment must meet to be in the best interests of the corporation one of rationality or reasonableness? The article concludes that the balance currently struck between a director's authority and accountability is weighted too heavily towards the former and, as a result, requires further reform.

I INTRODUCTION

The decision by Austin J in *Australian Securities and Investments Commission v Rich* (2009) 236 FLR 1 (*ASIC v Rich*) substantially redrew the contours of the business judgment rule in Australia. This article discusses the business judgment rule — and the decision in *ASIC v Rich* — through a focus on three main issues that are central to ensuring the striking of an effective balance between directors' authority and accountability, namely:

- Which party should bear the onus of proof in raising the defence?
- What is required for a director to be informed about the subject matter of the judgment?
- Is the standard that the director's judgment must meet to be in the best interests of the corporation one of rationality or reasonableness?

This article discusses these issues, and the struggle to strike an appropriate balance between competing policy concerns, with reference to the historical

* Associate Professor, Faculty of Law, University of New South Wales.

** Partner, Clayton Utz. The authors would like to thank Emily Rumble, Final Year Law Student, UNSW, and Charbel Azzi, Solicitor, Clayton Utz, for their research assistance.

development of the business judgment rule in Australia. This is supplemented by a review of the American Law Institute's ('ALI') *Principles of Corporate Governance*,¹ a key source for both the design of the business judgment rule in Australia, and Austin J's decision in *ASIC v Rich*.

II BACKGROUND

To understand the application of the business judgment rule in *ASIC v Rich*, and the current state of Australian law in this area, it is crucial to understand both the origins of the rule, and its application and treatment in the period between its introduction and *ASIC v Rich*.

A The Introduction of the Business Judgment Rule in Australia

Directors today are on the front line. So far as exposure to liability is concerned, the 'exposures' are increasing in both the criminal and the civil areas.²

The above words, quoted by the 1989 Cooney Report,³ clearly indicate the fear⁴ felt by Australian directors in regards to both criminal and civil liability at this time, due to what was perceived to be the subjective interpretation of the obligation on directors to exercise skill and care in the performance of their duties. The Cooney Report thus recommended the establishment of a clear objective duty of care for directors in Australian companies legislation.⁵ An objective standard was defined by the Cooney Report as 'one that all individuals would be expected to meet, regardless of their particular capacities and circumstances'.⁶

¹ ALI, *Principles of Corporate Governance: Analysis and Recommendations* (American Law Institute Publishers, rev ed, 2002).

² Kenneth McPherson, 'Directors' Responsibilities — A Regulator's Perspective' (Address delivered to the Company Directors Seminar, Adelaide, 1 June 1989), quoted in Senate Standing Committee on Legal and Constitutional Affairs, Parliament of Australia, *Company Directors' Duties* (1989) 27 [3.22] ('Cooney Report').

³ Cooney Report, above n 2.

⁴ This fear was compounded by the fact that the only common law protection afforded to directors in regards to their liability for business decisions was judicial reticence regarding the review of business judgments. See, eg, *Harlowe's Nominees Pty Ltd v Woodside (Lakes Entrance) Oil Co NL* (1968) 121 CLR 483, 493:

[d]irectors in whom are vested the right and duty of deciding where the company's interests lie and how they are to be served may be concerned with a wide range of practical considerations, and their judgment, if exercised in good faith and not for irrelevant purposes, is not open to review in the courts.

And see: *Howard Smith Ltd v Ampol Petroleum Ltd* [1974] AC 821, 832: 'courts of law [will not] assume to act as a kind of supervisory board over decisions within the powers of management honestly arrived at'.

⁵ The Cooney Report found that although s 229 of the *Companies Code* (the then principal statutory provision which governed the duties of company officers) on its face imposed an objective standard, the test being applied by the courts was a subjective one.

⁶ Cooney Report, above n 2, 25 [3.19].

In fleshing out this objective standard, the Cooney Committee considered adopting a business judgment rule similar to that developed by courts in the United States that provided special protection to directors' informed business decisions.⁷ The policy underlying a business judgment rule is said to be to recognise the need for directors to engage in considered risk taking and to protect the directors when those risks are part of an informed business judgment.⁸ The Cooney Committee recommended that a business judgment rule be introduced into Australian company law, that it should oblige directors to inform themselves of matters relevant to the administration of the company, and that directors 'should be required to exercise an active discretion in the relevant matter or, alternatively, to show a reasonable degree of care in the circumstances'.⁹

The federal government responded to the recommendations of the Cooney Committee by amending the Corporations Law¹⁰ to 're-enforce that the duty of care is an objective one, requiring a company officer to exercise the degree of care and diligence that a reasonable person in a like position as the officer of a corporation would exercise in the corporation's circumstances'.¹¹ However, these amendments to the Corporations Law did not include the introduction of a statutory business judgment rule, prompting Senator Hill, then in opposition, at the Second Reading of the Bill to say:

The Coalition would have preferred that the Bill also introduce a statutory business judgment rule. The advantage of such a rule is that it provides certainty for directors. Corporate law should encourage, and afford broad protection to, informed business judgments in order to stimulate risk taking and innovation and to give directors the confidence to make commercial decisions on their true merits.¹²

⁷ Ibid 29–30 [3.30].

⁸ See, eg, *ASIC v Maxwell* (2006) 59 ACSR 373; *ASIC v Macdonald [No 11]* (2009) 256 ALR 199, 245 [236]; *ASIC v Rich* (2009) 236 FLR 1, 128–9.

⁹ Cooney Report, above n 2, 31 [3.35] (emphasis added). The Cooney Report observed: American courts have developed a 'business judgment rule' which provides special protection to directors informed business decisions. The American Law Institute has devised a relatively precise formulation which is consistent with the rule developed by the courts which avoids much of the confusion that has arisen from the various ways in which the courts have started to rule. The main feature of the rule that the American Law Institute proposes is that a 'safe harbour' is created for a director or officer who makes a business judgment in good faith if: (a) he or she has no personal interest in the subject of the business judgment rule; (b) he or she is informed to an appropriate extent about the subject of the business judgment; and (c) he or she rationally believes that the business judgment is in the best interests of the company.

¹⁰ This is a reference to the coordinated legislation of 1990 whereby the Commonwealth and the states sought to produce a national system of regulation by each enacting the same legislation. It continued until 2001 when the states referred their powers to the Commonwealth. The relevant Acts were the *Corporations Act 1989* (Cth), with each state and territory enacting separate legislation: see, eg, *Corporations (Victoria) Act 1990* (Vic).

¹¹ Explanatory Memorandum, Corporate Law Reform Bill 1992 (Cth) 23 [82]. The new s 232(4) became effective on 1 February 1993.

¹² Commonwealth, *Parliamentary Debates*, Senate, 17 December 1992, 5297 (Robert Hill).

In 1997, the new federal government released its Corporate Law Economic Reform Program ('CLERP') policy papers. One of the areas proposed for reform was directors' duties. The then Treasurer, Mr Costello, spoke of the perceived difficulties in finding an appropriate balance between encouraging risk taking and ensuring the protection of investors. Prominent amongst the issues raised in the CLERP policy papers was the appropriateness of a proposal to develop a statutory business judgment rule.

The Explanatory Memorandum to the CLERP Bill 1998 stated that 'the fundamental purpose of the business judgment rule is to protect the authority of directors in the exercise of their duties, not to insulate directors from liability'.¹³ The Explanatory Memorandum also stated that the proposed statutory formulation of the business judgment rule would 'clarify and confirm the common law position that the Courts will rarely review bona fide business decisions'.¹⁴ Furthermore, under the proposed CLERP legislation, provided directors fulfilled their requirements, they would have an explicit safe harbour for any breach of their duty of care and diligence, and the merits of their business judgments would not be the subject of review by the Courts. The business judgment rule would therefore act as a rebuttable presumption in favour of directors which, if rebutted, would still require a plaintiff to establish that the directors breached their duty of care and diligence.¹⁵

The business judgment rule anticipated by the CLERP policy papers was inserted into the Corporations Law¹⁶ in March 2000. It provided:

180(2) — Business Judgment Rule

A director or other officer of a corporation who makes a business judgment is taken to meet the requirements of subsection (1), and their equivalent duties at common law and in equity, in respect of the judgment if they:

- (a) make the judgment in good faith for a proper purpose; and
- (b) do not have a material personal interest in the subject matter of the judgment; and
- (c) inform themselves about the proper subject matter of the judgment to the extent they reasonably believe to be appropriate; and
- (d) rationally believe that the judgment is in the best interests of the corporation.

The director's or officer's belief that the judgment is in the best interests of the corporation is a rational one unless the belief is one that no reasonable person in their position would hold.¹⁷

¹³ See Explanatory Memorandum, Corporate Law Economic Reform Program Bill 1998 (Cth) 17 [6.3].

¹⁴ Ibid 17 [6.4].

¹⁵ Ibid 18 [6.9].

¹⁶ See above n 10.

¹⁷ *Corporate Law Economic Reform Program Act 1999* (Cth) sch 1 (13 March 2000), inserting the provision at s 180(2).

Notwithstanding the introduction of a statutory business judgment rule in Australia, it is fair to say that, to date, it has not provided the certainty for directors which Senator Hill spoke of in such an aspiring manner in 1992.

*B From Enactment to ASIC v Rich: The Business Judgment Rule
Between 2000–09*

In the period between the introduction of the business judgment rule in 2000 and the case of *ASIC v Rich* in 2009, the business judgment rule had not been successfully invoked.¹⁸ However, the Australian Securities and Investments Commission's ('ASIC') high profile prosecution of former HIH Insurance Ltd directors Rodney Adler, Raymond Williams and Dominic Fodera saw an unsuccessful attempt to invoke the business judgment defence, as the NSW Supreme Court found that no actual decision had been made that could attract the operation of the rule.¹⁹

This failure to successfully invoke the business judgement rule was accompanied by criticism of the scope of the rule, most notably in the June 2007 submission to Treasury by the Australian Institute of Company Directors ('AICD'). The AICD argued that senior, experienced, and potential directors, particularly recently retired senior executives, were shying away from board positions especially in the listed environment, because of concerns about personal liability risks and damaged reputations arising from claims against them. Further, many directors were thought by the AICD to be forming the view that there was no longer a fair balance between risk and reward, or that the risks were exceeding the rewards of directorships.²⁰

In November 2007 in a speech to AICD members, the then-Chairman of ASIC Mr Tony D'Aloisio appeared to throw his weight behind the AICD view. He stated that 'it may be time ... to assess this balance between ensuring our boards take risks (so that our economy keeps growing) with protection of shareholders and creditors and consumers where individual liability may be appropriate'.²¹

In 2008, Professor Robert Baxt also indicated his support for the AICD position, and joined calls for law reform initiatives to relax director duties and encourage entrepreneurial spirit. Professor Baxt recommended governmental consideration of

¹⁸ H A J Ford, R Austin and I Ramsay, LexisNexis, *Ford's Principles of Corporations Law* (at August 2013) [8.310].

¹⁹ *ASIC v Adler* (2002) 41 ACSR 72, 173, 175–176, 196. See also *Gold Ribbon (Accountants) Pty Ltd (in liq) v Sheers* [2006] QCA 335, [186], [247].

²⁰ AICD, Submission to Corporations and Financial Services Division of The Treasury, *Review of Sanctions for Breaches of Corporate Law* (June 2007) 3. See also Australian Government, *Review of Sanctions in Corporate Law* (March 2007).

²¹ Tony D'Aloisio, 'An Update on ASIC's Priorities for 2007/2008 and How These Relate to AICD Members' (Paper presented at the AICD Luncheon, Sydney, 26 November 2007) <<http://www.asic.gov.au>>.

a new and improved business judgment rule, which would apply not only to s 180 of the *Corporations Act* but also to other provisions such as insolvent trading.²²

However, not all commentators writing in this period believed that a more extensive business judgment rule was necessary. The most prominent criticism was that a revamped business judgment rule would remain under-used as it would continue to require the defence to disprove the elements contained in the directors' duties provisions with the burden of proof shifting to the director.²³ For example, Neil Young QC noted that directors already have recourse to s 1318 of the *Corporations Act*, whereby if a director demonstrates to the Court that he or she has acted honestly, and in the circumstances ought fairly to be excused, the director may be excused by the court from civil liability for negligence, default, breach of trust, or breach of duty. Young cited the decision of Palmer J in *Hall v Poolman*²⁴ as demonstrating the capacity of courts to excuse directors from liability when they make honest mistakes.²⁵

²² Robert Baxt, 'Forgiving Directors for Insolvent Trading — Some Promising Signs' (February 2008) *Company Director Magazine* 48. The Federal Government subsequently issued the discussion paper: Corporations and Financial Services Division, The Treasury, *Insolvent Trading: A Safe Harbour for Reorganisation Attempts Outside of External Administration*, Discussion Paper (2010) which sought comment on a proposal to enact a business judgment rule in respect of the director's duty to avoid insolvent trading.

²³ Neil Young, 'Has Directors' Liability Gone Too Far or Not Far Enough? A Review of the Standard of Conduct Required of Directors Under Sections 180 to 184 of the *Corporations Act*' (2008) 26(4) *Companies and Securities Law Journal* 216, 223.

²⁴ (2007) 215 FLR 243. In *Hall v Poolman*, Palmer J found that a director's contravention of s 588G(2) should be wholly relieved pursuant to ss 1317S and 1318 for a specified period of time. See also *Re McLellan; The Stake Man Pty Ltd v Carroll* (2009) 76 ACSR 67, excusing pursuant to s 1317S(2) a sole director of a company from contravention of the insolvent trading provisions.

²⁵ Young, above n 23, 222–3. However, it is worth noting the disadvantages of seeking relief under s 1318, including the uncertainty of obtaining the desired outcome, given judicial discretion, and the differing consequences for a director's reputation to seeking relief from liability under s 1318 rather than using the business judgment rule defence to prevent a finding of breach of duty. Furthermore, a survey of the use of s 1318 indicates its limited success in Australian law, thus reducing its attractiveness: see Steven Wong, 'Forgiving a Director's Breach of Duty: A Review of Recent Decisions' (Research Report, Centre for Corporate Law and Securities Regulation, 1 December 2008) 6 <http://www.law.unimelb.edu.au/files/dmfile/stevenwong_essay_6_May_20091.pdf>. See also Jason Harris, 'Relief from Liability for Company Directors: Recent Developments and Their Implications' (2008) 12 *University of Western Sydney Law Review* 152. Most recently, the case of *ASIC v Healey [No 2]* [2011] FCA 1003 (31 August 2011) demonstrated judicial reluctance to grant relief under s 1317S or s 1318 once directors have been found to have breached their duty. Justice Middleton held that 'the seriousness of the contraventions preclude[d] [him] from exercising discretion to relieve any of the defendants from liability, having regard to the principles of general deterrence': at [133].

Subsequently, Young expressed the view that directors should be held to ‘a fairly high standard of reasonable care and diligence’ as the company board is the responsible corporate organ for managing the affairs of the company, and the law should aim to set norms of behaviour that will deter unwanted conduct.²⁶ Furthermore, Young has probed the utility of the business judgment rule in s 180(2), questioning whether a director who has made a decision in good faith and for a proper purpose, has properly informed him or herself, and the decision is rational, could nevertheless be in breach of s 180(1).²⁷

It was in this context and against this background of criticism that the decision of Austin J in *ASIC v Rich* was handed down on 18 November 2009. However, before turning to the findings in the judgment, it is important to review the experiences of the United States in applying the business judgment rule, given the statements of many commentators, as well as Austin J himself, that the Australian business judgment rule in s 180(2) was largely drawn from the ALI’s statement of the business judgment rule.²⁸

III THE ALI’S BUSINESS JUDGMENT RULE

The business judgment rule has existed in American corporate law for almost two centuries.²⁹ However, in 1992, the ALI sought to formulate a standard business judgment rule based on that historical development set out in para 4.01(c) of its *Principles of Corporate Governance*.³⁰

The ALI’s business judgment rule is based largely on two rationales. First, it protects directors from shareholders second guessing directors’ decisions with the benefit of hindsight. In line with Australian concerns, American commentators fear that directors might not take calculated risks, or even serve on boards, if they are open to derivative litigation as a result of every failed board decision. Secondly, it has

²⁶ Neil Young, ‘Directors Duty of Care and Diligence: A Review in Light of the Recent Decision in *ASIC v MacDonald (No 11)*’ in R P Austin and A Y Bilski (eds), *Directors in Troubled Times* (Ross Parsons Centre of Commercial, Corporate and Taxation Law, 2009) 94.

²⁷ *Ibid* 81.

²⁸ *ASIC v Rich* (2009) 236 FLR 1, 146 [7256]–[7257]; Ford, Austin and Ramsay, above n 18, [8.310]; Cooney Report, above n 2, 31 [3.35]; Companies and Securities Law Review Committee, *Company Directors and Officers: Indemnification, Relief and Insurance*, Discussion Paper No 9 (1990) 32 [69].

²⁹ S Samuel Arsht, ‘The Business Judgment Rule Revisited’ (1979) 8 *Hofstra Law Review* 93; *Percy v Millaudon*, 8 Mart (ns) 68, 77–78 (La, 1829); *Godbold v Branch Bank at Mobile*, 11 Ala 191, 199 (1847).

³⁰ ALI, above n 1, 166. ‘The formulation of the business judgment rule set forth in paragraph 4.01(c) is believed to be consistent with present law as it would be interpreted in most jurisdictions today, and each of the rule’s basic elements (paragraph 4.01(c)(1)–(3)) is supported by substantial precedential authority’; R P Austin, H A J Ford and I M Ramsay, *Company Directors: Principles of Law and Corporate Governance* (LexisNexis Butterworths, 2005) 242–3 [6.12].

been said that directors are better suited than courts to make business decisions, and thus the rules protect directors from liability for honest mistakes of judgment or unpopular business decisions.³¹

Paragraph 4.01(c) states that:

A director or officer who makes a business judgment in good faith fulfils the duty [of care] ... if the director or officer:

- 1 is not interested in the subject of the business judgment;
- 2 is informed with respect to the subject of the business judgment to the extent the director or officer reasonably believes to be appropriate under the circumstances; and
- 3 rationally believes that the business judgment is in the best interests of the corporation.³²

A comparison with s 180(2) demonstrates a close resemblance to the Australian business judgment rule, except that the ALI's version does not have the 'qualification' that follows s 180(2)(d) regarding a view that 'no reasonable person' would possess. Further, para 4.01(d) provides that a person challenging the conduct of a director or officer under this section has the burden of proving a breach of the duty of care, including the inapplicability of the provisions as to the fulfilment of duty under para 4.01(c), and, in a damages action, the burden of proving that the breach was the legal cause of damage suffered by the corporation.³³

A Prerequisite of a Conscious Exercise of Judgment

The ALI states that to be afforded the protection of the business judgment rule, a decision must have been consciously made and judgment must, in fact, have been exercised. The Delaware case of *Aronson v Lewis*,³⁴ consistent with the ALI's formulation, held that

³¹ ALI, above n 1, 130–1; Ann Scarlett, 'Confusion and Unpredictability in Shareholder Derivative Litigation: The Delaware Court's Response to Recent Corporate Scandals' (2008) 60 *Florida Law Review* 589, 600.

³² ALI, above n 1, 134.

³³ *Ibid.* See also Austin, Ford and Ramsay, above n 30, 238–9 [6.7]. The ALI formulation is not the only way in which the business judgment rule is expressed in the US, but it is the formulation that was relied on for the drafting of s 180(2). For another formulation, see, eg, *Robotti & Co LLC v Liddell*, 2010 Del Ch LEXIS 4, 46–7:

The business judgment rule, as a general matter, protects directors from liability for their decisions so long as there exist 'a business decision, disinterestedness and independence, due care, good faith and no abuse of discretion and a challenged decision does not constitute fraud, illegality, ultra vires conduct or waste'. There is a presumption that directors have acted in accordance with each of these elements, and this presumption cannot be overcome unless the complaint pleads specific facts demonstrating otherwise. Put another way, under the business judgment rule, the Court will not invalidate a board's decision or question its reasonableness, so long as its decision can be attributed to a rational business purpose.

³⁴ 473 A 2d 805 (Del, 1984).

‘the business judgment rule operates only in the context of director action ... it has no role where directors either abdicated their functions, or absent a conscious decision, failed to act’.³⁵ Further, *Kaplan v Centex Corp* noted that it must be shown ‘that [the] director[’s] judgment was brought to bear with specificity on the transactions’.³⁶

In other words, the business judgment rule will not protect directors in situations where no business decision-making has occurred, including those situations where directors have failed to contemplate the need for a decision to be made.³⁷

B Prerequisite of Good Faith and No Interest

Good faith and disinterested decision-making are prerequisites for the application of the business judgment rule’s safe harbour.³⁸ Good faith has been interpreted to mean that a decision is motivated by acting in the corporation’s best interest, whilst disinterested decision-making means that the director does not have a conflict of interest.³⁹ To render the business judgment rule inapplicable, the party challenging a decision has the onus of establishing a lack of good faith and disinterested decision-making. If this challenge is successful, then the ‘burden shifts to the director to prove that the transaction was found reasonable to the corporation’.⁴⁰

There have been some suggestions that a good faith standard alone should suffice for a director’s actions to be protected. However, the ALI maintains that courts have articulated that a mere good faith test may, depending on the court’s interpretation of that test, provide too much insulation for directors and officers.⁴¹

C Prerequisite of an Informed Decision

Directors are required to inform themselves prior to making the business decision of all material information reasonably available to them.⁴² This informed decision

³⁵ Ibid 813.

³⁶ *Kaplan v Centex Corp*, 284 A 2d 119, 124 (Del Ch, 1971).

³⁷ ALI, above n 1, 167–8. This includes situations involving omissions. For example, the ALI offers the example of directors failing to even consider the need for an effective audit process to prevent executives from absconding with corporate funds as a scenario where para 4.01(c) will not protect directors — as the directors in question will have failed to make a decision attracting the application of the business judgment rule.

³⁸ Ibid 169.

³⁹ Dennis J Block, Nancy E Barton and Stephen A Radin, *The Business Judgment Rule: Fiduciary Duties of Corporate Directors* (Prentice Hall Law & Business, 3rd ed, 1989) 13–17, 19.

⁴⁰ ALI, above n 1, 169–170, citing *Treadway Companies Inc v Care Corp*, 638 F 2d 357, 382 (2nd Cir, 1980).

⁴¹ Ibid 174.

⁴² Ibid 170. See also *Casey v Woodruff*, 49 NYS 2d 625, 643 (Sup Ct, 1944); *Miller v American Telephone & Telegraph Co*, 507 F 2d 759, 762 (3rd Cir, 1974); *Treadway Companies Inc v Carecorp*, 638 F 2d 357, 384 (2nd Cir, 1980).

prerequisite focuses on the preparedness of a director or officer in making a business decision, as opposed to the quality of the decision itself.⁴³

The ALI lists the following factors that may have to be taken into account in judging a director's reasonable belief about what was 'appropriate under the circumstances':

- (a) the importance of the business judgment to be made;
- (b) the time available for obtaining information and deciding the extent to which he or she should be informed;
- (c) the costs related to obtaining information;
- (d) the director's confidence in those who explored a matter and those making presentations;
- (e) the state of the corporation's business at the time and the nature of competing demands for the board's attention;
- (f) the different backgrounds and experience of individual directors, the distinct role each plays in the corporation, and the general value of maintaining board cohesiveness; and
- (g) the general views or specialised experience of colleagues.⁴⁴

The classic American example of failing to make an informed decision is encapsulated in *Smith v Van Gorkom*,⁴⁵ where the board approved a merger after only two hours of consideration, relying on a 20 minute presentation by the CEO who had arranged the merger, and despite the advice of senior management that the offer was too low.

D The 'Rationally Believes' Requirement

The ALI states that if the requirements of 'good faith' and para 4.01(c)(1)–(2) are met, the business judgment rule will protect a director or officer from liability for a business judgment provided the director or officer actually believes that the business judgment is in the best interests of the corporation, and that belief is rational.⁴⁶ This rational belief test is the ALI's amalgam of various formulations used in various US jurisdictions, such as 'egregious', 'reckless', and 'irrational', to denote a decision which no person of ordinary sound business judgment would make.⁴⁷

In ascertaining the ALI's perspective on how this test should be interpreted, the ALI describes Delaware case law as establishing that 'gross negligence is the standard to

⁴³ ALI, above n 1, 170.

⁴⁴ Ibid 171.

⁴⁵ 488 A 2d 858 (Del, 1985).

⁴⁶ ALI, above n 1, 172.

⁴⁷ Charles Hansen, 'The Duty of Care, The Business Judgment Rule and the American Law Institute Corporate Governance Project' (1993) 48 *Business Lawyer* 1355, 1375.

be applied in deciding ... whether the directors may be held liable for reaching the wrong decision',⁴⁸ and that the rational belief standard may be similar to an absence of gross negligence.⁴⁹ Conversely, Californian case law supports the view that a director or officer's business judgment must be 'reasonable' to be protected.⁵⁰

However, the ALI states that sound public policy dictates that directors and officers should be given greater protection than a 'reasonableness' test would afford.⁵¹ The ALI's policy is not clearly specified, but appears to be that the rational belief test provides a 'significantly wider range of discretion'⁵² than 'reasonable', whilst also accommodating some objective element, as seen in the Delaware 'gross negligence' test.⁵³

E Process v Substance Approach

The ALI draws a distinction between the level of judicial scrutiny of the director's decision itself, and review of the process the directors used to arrive at the decision. The ALI approach limits judicial scrutiny of the substance of the director's decision. However, the ALI formulates a standard of reasonable belief regarding the process the director uses to reach the decision, or, more specifically, whether the director gathered adequate information before acting. When it comes to the substance of the director's decision, however, the ALI's version of the business judgment rule lowers the standard of care to a rational belief.⁵⁴

F Onus of Proof

The ALI formulation places the burden of proving the failure of a director to comply with the duty of care, including the inapplicability of the business judgment rule in para 4.01(c), on the party challenging the director's conduct.⁵⁵

⁴⁸ ALI, above n 1, 173.

⁴⁹ Ibid 172. See also *Rabkin v Philip A Hunt Chemical Corp*, 547 A 2d 963 (Del Ch, 1986): 'In the corporate area, gross negligence would appear to mean "reckless indifference to or a deliberate disregard of the stockholders"' [quoting *Allaun v Consolidated Oil Co*, 147 A 257, 261 (Del Ch, 1929)], or actions which are "without the bounds of reason" [quoting *Gimbel v Signal Companies Inc*, 316 A 2d 599, 615 (Del Ch, 1974)]; *McPadden v Sidhu*, 964 A 2d 1262, 1274 (Del Ch, 2008); Franklin Gervurtz, 'The Business Judgment Rule: Meaningless Verbiage or Misguided Notion' (1994) 67 *Southern California Law Review* 287, 301.

⁵⁰ See, eg, ALI, above n 1, 174. See also *McDonnell v American Leduc Petroleum Ltd*, 491 F 2d 380, 384 (2nd Cir, 1974), in which the Second Circuit Court of Appeals applied Californian precedent and concluded that the 'business judgment rule protects only reasonable acts of a director or officer'.

⁵¹ ALI, above n 1, 173, 178–9.

⁵² Ibid 136.

⁵³ Ibid 173.

⁵⁴ Gervurtz, above n 49, 300–1.

⁵⁵ ALI, above n 1, 179.

Courts within the American jurisdiction of Delaware have described the rule as a presumption that, in making a business decision, the directors of a corporation acted on an informed basis, in good faith, and in the honest belief that the action was taken in the best interests of the company.⁵⁶ The ALI declines to use this phraseology, but expects the outcome to be the same.⁵⁷

The party alleging liability may rebut the presumption by showing either:

- that the directors violated any one of their fiduciary duties; or
- that the business judgment rule is inapplicable because the directors committed an act of fraud, illegality or waste.⁵⁸

If the party alleging liability fails to rebut the presumption, then the business judgment rule will protect the directors from liability for their decision.⁵⁹ If the party alleging liability rebuts the presumption, the director must prove that the challenged transaction was ‘entirely fair’ to the corporation and its shareholders.⁶⁰

IV *ASIC v RICH*

Until its collapse in May 2001, OneTel Ltd (‘OneTel’) was a large listed telecommunications company in Australia, which counted Publishing and Broadcasting Limited (‘PBL’) and News Corporation (‘News Corp’) as two of its major shareholders. In April 2001, there were ultimately futile discussions about the possibility of PBL and News Corp injecting cash into OneTel through a rights issue. In May 2001, PBL and News Corp united to remove CEO Jodie Rich and finance director Mark Silbermann from the OneTel board. The board then appointed voluntary administrators to OneTel.⁶¹

Proceedings were commenced by ASIC against four defendants, but orders were made before the hearing in respect of two of those defendants, so that the

⁵⁶ Deborah De Mott, ‘Inside the Corporate Veil: The Character and Consequences of Executive Duties’ (2006) 19 *Australian Journal of Corporate Law* 251, 261; *Aronson v Lewis*, 473 A 2d 805, 812 (Del, 1984); *Carsanaro v Bloodhound Technology Inc*, 65 A 3d 618, 637 (Del Ch, 2013):

[b]ecause the business judgment rule establishes a presumption in favor of the directors, a plaintiff only can proceed by alleging facts sufficient to overcome one of the elements of the rule. Under the business judgment rule, the burden of pleading and proof is on the party challenging the decision.

⁵⁷ ALI, above n 1, 178–80.

⁵⁸ Ann Scarlett, ‘A Better Approach for Balancing Authority and Accountability in Shareholder Derivative Litigation’ (2008–09) 57 *Kansas Law Review* 39, 59–60.

⁵⁹ *Ibid.*

⁶⁰ Scarlett, above n 31, 600.

⁶¹ *ASIC v Rich* (2009) 236 FLR 1, 16 [8], 206 [7515].

case proceeded against only Rich and Silbermann. In this case ASIC sought civil penalties for alleged breaches of the statutory duty of care and diligence arising out of the collapse of OneTel in May 2001.

Justice Austin's introduction to the judgment⁶² shows that ASIC sought to establish, as the central component of its case, that the financial position of the OneTel Group was much worse over the relevant period than the information provided to the board revealed, and that the forecast that had been provided to the board had no proper basis. ASIC also contended, amongst other things, that the defendants were aware of OneTel's poor financial position, or ought to have been aware, and failed to make proper disclosure to the board. The hearing lasted 232 days.

Ultimately, Austin J held that ASIC had failed to prove its case that Mr Rich and Mr Silbermann had contravened s 180.⁶³ At the heart of this decision was his Honour's conclusion that the rational belief required as an element of the business judgment rule in s 180(2) does not have to be reasonable. In other words, a director can invoke the business judgment rule if he shows that he arrived at the business judgment after a reasoning process 'whether or not the reasoning process was convincing to the judge and therefore reasonable in an objective sense'.⁶⁴

A Who Bears the Onus of Proof?

The starting point for the application of the business judgment rule must be identifying the party bearing the onus of proving its elements. This is particularly true given the 'substantial findings of fact' often required in cases involving the business judgment rule, which Austin J described as ensuring that the issue of onus of proof was 'critically important'.⁶⁵

In construing the provision, Austin J noted that the statutory provision itself does not expressly indicate which party carries the onus of proving the application of the business judgment rule, and that he could not 'extract any reliable indication either way from simply reading the text'.⁶⁶ Nor did his Honour find any significant assistance from the statement of legislative policy considerations in the Explanatory Memorandum or Second Reading Speech.⁶⁷ His Honour then moved to consider the formulation of the United States business judgment rule upon which, to a large degree, s 180(2) was modelled. Whilst the ALI's business judgment rule appears

⁶² Ibid 184–5 [7391].

⁶³ Ibid.

⁶⁴ Ibid 154 [7289].

⁶⁵ Ibid 146 [7259].

⁶⁶ Ibid 147–8 [7264].

⁶⁷ Ibid 148–9 [7265]–[7268].

to place the onus upon the plaintiff,⁶⁸ his Honour considered that if Parliament had intended to replicate that position, it might have more clearly done so.

In ascertaining where the onus of proof should lie, Austin J discussed the conflicting practical consequences emerging from placing the onus of proof on the directors and officers or on the party asserting liability — in this case, ASIC. In particular, his Honour focused on the trade off between legislative attempts to encourage responsible risk taking, which he believed would be hampered by placing the onus of proof on directors, against placing a higher burden on those seeking to hold directors against a higher level of accountability. In particular, he believed that the latter would ‘add substantial elements to the burden of proof of contravention that were not present in previous statutory formulations of the duty of care and diligence’.⁶⁹

Justice Austin ultimately took the stance that the question of whether the plaintiff or defendant bore the onus of proof was ‘an important one that will eventually need to be resolved at an appellate level’, but that, although bearing in mind the US approach, s 180(2) casts the onus on the defendants.⁷⁰

His Honour based this conclusion on two reasons. First, he concluded that placing the onus on the plaintiffs would have been inconsistent with the clear intention expressed in the Explanatory Memorandum and the Second Reading Speech that there was to be no reduction in the statutory requirement. Secondly, he found that if the plaintiff was required to establish that the four criteria had not been met, then as part of that obligation the plaintiff would be required to show that the defendant’s business judgment was not made in good faith for a proper purpose. Proving this would amount to a requirement to prove a more serious contravention of the law — a contravention of s 181. His Honour also noted that the four elements, required to

⁶⁸ See *ibid* 147 [7261], citing Deborah De Mott, ‘Legislating Business Judgment — A Comment from the United States’ (1998) 16 *Company and Securities Law Journal* 575, 576:

As defined by cases in Delaware (by far the United States jurisdiction with the most fully developed body of corporate law doctrine), the business judgment rule clearly creates a presumption on behalf of directors that the prerequisites to the rule’s applicability have been satisfied [citing, for example, *Aronson v Lewis*, 473 A 2d 805 (Del, 1984)]. As a consequence, the plaintiff bears the onus as to whether the director’s decision was disinterested, made in good faith with adequate information, and made with a rational belief that the decision served the corporation’s best interests.

⁶⁹ *ASIC v Rich* (2009) 236 FLR 1, 148 [7266].

⁷⁰ *Ibid* 149 [7269]. See also Young, above n 26, 80, who states that the business judgment rule:

does not attempt to define the standard of care required of directors making business judgment nor is it formulated as a preliminary presumption that explicitly puts the burden of rebuttal on the party alleging a breach of section 180(1). Rather it is framed as a deemed provision that will operate only if the four conditions in sub-section (2) are proved. It is hard to see how a director can avail himself or herself of the business judgment rule without adducing evidence satisfying these conditions.

be proven to make good the business judgment defence, were matters principally within the knowledge of the director or officer.⁷¹

Nevertheless, Austin J appears to have been troubled that the view that he ultimately came to was inconsistent with the objectives of the Explanatory Memorandum to the CLERP Bill of 1998. This Explanatory Memorandum stated that the purpose of the CLERP legislation was to protect directors exercising their duties from review by the courts, and not to discourage them from taking advantage of opportunities involving responsible risk taking. His Honour reconciled his position on the onus issue through a conclusion that placing the onus of proof on directors was not necessarily incompatible with the Explanatory Memorandum's purpose, as

[i]t may happen in practice that the evidential burden can be shifted to the plaintiff relatively easily, if the defendant addresses the statutory elements in his or her affidavit, though the price to be paid is that the defendant is exposed to cross-examination on those matters.⁷²

Justice Austin's concerns were well-founded. The placement of the onus of proof on the director ensures that the elements of the director's decision — contrary to the very purpose of a business judgment rule — are subject to judicial scrutiny. A business judgment rule is believed desirable because judges lack the information and skill necessary to evaluate business judgments, business is inherently risky, and the quality of a business decision cannot always be judged by the immediate results. Therefore, liability for a decision that produces bad results would discourage a board from entrepreneurial risk taking and make it difficult to secure the services of able and experienced corporate directors.⁷³

Equally, the business judgment rule is not meant to insulate directors from liability. Put another way, as argued by Stephen Bainbridge, the business judgment rule reflects the need to reach a compromise between recognising the director's authority and discretion to decide, and the need for those directors to be accountable.⁷⁴ Although in the interpretation of a provision of an Act, a construction that would promote the purpose or object underlying the Act is to be preferred to a construction that would not promote that purpose or object, this may be of little assistance where a statutory provision strikes a balance between competing policy concerns, as is the case here. Consequently, it is the text — construed according to such principles of

⁷¹ *ASIC v Rich* (2009) 236 FLR 1, 149 [7269].

⁷² *Ibid* 149–50 [7270].

⁷³ Randy Holland, 'Delaware Director's Fiduciary Duties: the Focus on Loyalty' (2009) 11 *University of Pennsylvania Journal of Business Law* 675, 679; William Meade Fletcher et al, *Fletcher Cyclopedia of the Law of Corporations* (Thomson/West, 1931) 1037.

⁷⁴ Stephen Bainbridge, 'The Business Judgment Rule as Abstention Doctrine' (2004) 57 *Vanderbilt Law Review* 83, 84.

interpretation as provide rational assistance in the circumstances of the particular case — that must control the court’s interpretation of the provision.⁷⁵

That is not to say that Austin J’s construction of the legislation is inconsistent with the text of the provision. Conversely, it is for Parliament to revisit the business judgment rule and clarify the question of onus. It should be noted that a change in onus would not require the plaintiff to establish that all four criteria have not been met, but rather, that where one of the criteria has not been established, the safe harbour is not available. If one of the criteria is not met, then the Court will examine whether s 180(1) has been complied with. It is to be expected that in most cases alleging negligence the focus would be on s 180(2)(c)–(d), as was the case in *ASIC v Rich*.

B *What is a Business Judgment?*

Section 180(3) of the *Corporations Act* provides that for the purposes of s 180:

‘business judgment’ means any decision to take or not take action in respect of a matter relevant to the business operations of the corporation.

Justice Austin, drawing on the experiences of the United States, suggests that, in general, decisions to enter into transactions for financial purposes are business judgments. The question for his Honour in the proceedings was thus ‘how much further the concept of business judgment is extended into the realm of management, organisation and planning’.⁷⁶

His Honour found that to be a business judgment for the purposes of the Australian definition, there had to be a decision to take or not take action in respect of matters relevant to the business operations of the corporation. The important element, according to his Honour, is the need for a conscious decision — that is to say, whether or not the directors or officers in fact turned their minds to the matter.⁷⁷

Justice Austin also noted that in using the language ‘in respect of a matter relevant’, s 180(3) utilises an expression well recognised as being of great breadth. Accordingly, he thought it inevitable that the Courts would give the definition of ‘business judgment’ a wide interpretation.⁷⁸ He concluded that the following decisions would be covered by the business judgment rule:

- decisions that are preparatory to the making of a business decision;
- decisions relating to corporate personnel;

⁷⁵ *Carr v Western Australia* (2007) 232 CLR 138, 143; *Alcan (NT) Alumina Pty Ltd v Commissioner of Territory Revenue* (2009) 239 CLR 27. See also Murray Gleeson, ‘The Meaning of Legislation: Context, Purpose and Respect for Fundamental Rights’ (2009) 20 *Public Law Review* 26, 32–3.

⁷⁶ *ASIC v Rich* (2009) 236 FLR 1, 150 [7272].

⁷⁷ *Ibid* 151 [7277].

⁷⁸ *Ibid* 151 [7276].

- decisions relating to the termination of litigation;
- the setting of policy goals;
- the apportionment of responsibilities between the Board and senior management; and
- decisions about planning, budgeting and forecasting.⁷⁹

Notwithstanding the breadth of the statutory language, not all action taken by directors and officers will fall within the definition of a business judgment. Because of the need for directors or officers to make a conscious decision (that is to actively turn their minds to the matter), the discharge by them of their ‘oversight’ responsibilities (that is monitoring the company’s affairs and policies and maintaining familiarity with its financial position) is not covered. On his Honour’s interpretation, this does not involve any business judgment as defined.⁸⁰ More recently, the Full Federal Court in *Fortescue Metals*⁸¹ adopted a similar approach in relation to compliance with the continuous disclosure obligations under the *Corporations Act*, stating that

the decision not to disclose the true effect of the agreements cannot be described as ‘business judgment’ at all. A decision not to make accurate disclosure of the terms of a major contract is not a decision related to the ‘business operations’ of the corporation. Rather it is a decision related to compliance with the requirements of the Act.⁸²

Justice Austin thus found, generally speaking, that the case before him was not one where the defendant directors had failed to turn their minds to decisions that ASIC alleged they should have taken. Rather, to a substantial degree, it was a case where the defendants considered the matters of which ASIC complained and made decisions with which ASIC disagreed. In those circumstances, the decisions by the defendant directors taken in planning, budgeting and forecasting were capable of receiving the protection of the Australian statutory business judgment rule.⁸³

⁷⁹ Ibid 150 [7273]–[7274]; Paul Redmond, ‘Safe Harbours or Sleepy Hollows: Does Australia need a Statutory Business Judgment Rule?’ in Ian Ramsay (ed), *Corporate Governance and the Duties of Company Directors* (University of Melbourne, 1997) 195.

⁸⁰ *ASIC v Rich* (2009) 236 FLR 1, 151 [7278].

⁸¹ *ASIC v Fortescue Metals Group Ltd* (2011) 190 FCR 364.

⁸² Ibid 427 [197] (Keane CJ). The High Court of Australia overturned the Full Court’s judgment in *Forrest v ASIC* (2012) 291 ALR 399, but did not need to address the business judgment rule so that Keane CJ’s observations remain good law.

⁸³ *ASIC v Rich* (2009) 236 FLR 1, 150 [7274].

C Good Faith for a Proper Purpose and No Material Personal Interest

The requirement in s 180(2)(a)–(b) of making the judgment ‘in good faith for a proper purpose’ and not having a ‘material personal interest’ in the subject matter of the business judgment did not require detailed examination in *ASIC v Rich*.⁸⁴

ASIC submitted to his Honour in relation to the former that it was doubtful whether the element of good faith could be established if the directors or officers had failed to take action, in discharge of their responsibility to supervise, and the inaction has resulted in a failure to discover substantial corporate losses. His Honour said in response to this submission:

It seems to me that as far as the Australian law is concerned, the problem in such a case is that the officer who has failed to take action has probably not made any decision not to take action, and therefore there is no business judgment to be protected. If an officer decides not to do something, and that decision results in a failure to discover substantial corporate losses, there may well be a question of good faith on the facts, but there is no reason in principle why good faith cannot be established by evidence.⁸⁵

It was accepted by ASIC that the latter issue of material personal interest did not arise.⁸⁶

D Informing Oneself about the Subject Matter

Section 180(2)(c) requires that directors ‘inform themselves about the subject matter of the judgment to the extent they reasonably believe to be appropriate’.⁸⁷

Justice Austin endorsed ASIC’s list of factors said to be relevant to determining the reasonableness of the purported discharge of this element.⁸⁸ However, he rejected ASIC’s submission that regard should be had not only to what the director knew, but also what he should have known.⁸⁹

⁸⁴ For a discussion of these requirements, see Andrew Lumsden, ‘The Business Judgment Defence: Insights from *ASIC v Rich*’ (2010) 28 *Company and Securities Law Journal* 164, 171–4.

⁸⁵ *ASIC v Rich* (2009) 236 FLR 1, 152 [7281].

⁸⁶ *Ibid* 152 [7282].

⁸⁷ *Corporations Act 2001* (Cth) s 180(2)(c).

⁸⁸ *ASIC v Rich* (2009) 236 FLR 1, 152 [7283], listing the relevant factors as: (1) the importance of the business judgments to be made; (2) the time available for obtaining information; (3) the costs related to obtaining information; (4) the director or officer’s confidence in those exploring the matter; (5) the state of the company’s business at that time and the nature of competing demands on the board’s attention; (6) whether or not material information is reasonably available to the director.

⁸⁹ *Ibid* 152–3 [7284].

In his Honour's view, the statutory language related to the decision-making occasion rather than to the general state of knowledge of the director. It requires the director to become informed about the subject matter of the decision prior to making it. His Honour held that:

The qualifying words, 'to the extent they reasonably believe to be appropriate', convey the idea that protection may be available even if the director was not aware of the available information material to the decision, if he reasonably believed he had taken appropriate steps on the decision-making occasion to inform himself about the subject matter.⁹⁰

So much may be accepted. But what work does the word 'reasonably' perform if it is not an objective reality check for the belief held by the director or officer? Would the protection of which his Honour speaks be available if the directors' enquiries, although sufficient in their minds, were objectively inadequate?

In summarising the operation of the business judgment rule Austin J observed that the rule will operate only where:

- [the directors] make their decision after informing themselves about the subject matter to the extent they believe to be appropriate having regard to the practicalities listed at 23.9.5.4; [and]
- their belief about the appropriate extent of information gathering is reasonable in terms of the practicalities of the information gathering exercise (including such matters as the accessibility of information and the time available to collect it).⁹¹

The latter point suggests that there is some objective requirement about the information gathering process. However, this sits uncomfortably with his Honour's earlier analysis.

The ALI explained that US courts will not interfere in matters of business judgment if reasonable diligence and care have been exercised.⁹² The requirement in para 4.01(c)(2) of reasonable belief has both subjective and objective content; a reasonable belief will not exist if the director does not believe the inquiry made is adequate, or if the director is unreasonable in believing that the inquiry made is adequate.⁹³ While Austin J's reasoning seems to be weighted too heavily towards the subjective belief of the director and with insufficient regard for what is objectively adequate, the later summary seems to be consistent with the ALI position.

While the text of s 180(2)(c) refers to the director's belief, suggesting a subjective element, the requirement of reasonableness in Australia also suggests that there is

⁹⁰ Ibid.

⁹¹ Ibid 155 [7294].

⁹² ALI, above n 1, 170.

⁹³ Ibid 157, 171.

an objective test.⁹⁴ An objective element is warranted because it better balances the competing demands of recognising a director's need to make business decisions while also being accountable.

E Rational Belief

The fourth element of the business judgment rule in s 180(2) is that the director must 'rationally believe that the judgment is in the best interests of the corporation'.⁹⁵ This requirement is qualified by the statement that 'the director's or officer's belief that the judgment is in the best interests of the corporation is a rational one unless the belief is one that no reasonable person in their position would hold'.

ASIC submitted that the rationality of a director's belief is determined by whether it is reasonable. In that sense, the Australian provision differs from the rule that applies under the American principles. The Australian defence, if ASIC's submissions were supported, would not be available if the director's decision was based on an unreasonable belief.⁹⁶

His Honour noted that some 'unfortunate consequences' would flow from accepting ASIC's submission, including that:

- since the s 180(1) duty of care is based on reasonableness, the business judgment rule would have no work to do if it only applied to decisions that were reasonable; and
- ASIC's interpretation of the business judgment rule would differentiate our business judgment rule from that of the United States, which was not the intention of the drafters of the Australian legislation.⁹⁷

⁹⁴ See *Shafron v ASIC* (2012) 286 ALR 612, 617–618 [18]. See also *WA Pines Pty Ltd v Bannerman* (1980) 30 ALR 559, 572 (Lockhart J):

In my opinion the words in s 155 'has reason to believe ...' mean that the Commission must believe that a person is capable of furnishing information, producing documents or giving evidence; and there must be reasonable grounds or cause for that belief, before the powers conferred by s 155(1) may be exercised.

And see *Zecevic v DPP (Vic)* (1987) 162 CLR 645, 673 (Deane J) (finding the test for self defence that required reasonable belief was a mixture of the objective and subjective); *Lemoto v Able Technical Pty Ltd* (2005) 63 NSWLR 300, 332 [137] (finding that the solicitors state of mind of a reasonable belief under the *Legal Profession Act 1987* (NSW) had to be objective); *Healy v Commissioner for Consumer Protection* [2010] WASC 177 (22 July 2010) [35] (Heenan J) (holding that the reasonableness of the appellant's belief that they could supply a heating system was not dependent on the appellant's actual state of knowledge, but on what a reasonable belief in the circumstances would have been).

⁹⁵ *Corporations Act 2001* (Cth) s 180(2)(d).

⁹⁶ *ASIC v Rich* (2009) 236 FLR 1, 153 [7287].

⁹⁷ *Ibid* 153–4 [7288].

Justice Austin referred to the ALI's formulation of the business judgment rule and the Shorter Oxford English Dictionary to determine that the standard of the word 'rational' is less than that of 'reasonable'.⁹⁸ His Honour acknowledged that one meaning of the word 'rational' in the Shorter Oxford English Dictionary is 'agreeable to reason' or 'reasonable', but preferred another meaning: 'based on, derived from, reason or reasoning'.⁹⁹ His Honour thought it plausible to say that the drafters of the sub-section intended to capture the latter meaning so that:

the director's or officer's belief would be a rational one if it was based on reason or reasoning (whether or not the reasoning was convincing to the judge and therefore 'reasonable' in an objective sense), but it would not be a rational belief if there was no arguable reasoning process to support it. The drafters articulated the latter idea by using the words 'no reasonable person in their position would hold'.¹⁰⁰

Therefore, a director's belief that a certain decision is in the best interests of the corporation will be rational, on his Honour's view, if there is some arguable reasoning process to support it. This is so whether or not no reasonable person in the director's position would share that opinion. His Honour's preferred construction of sub-s (d) interprets the Australian business judgment rule in a way that is similar to, but more forgiving than, the ALI's interpretation. This construction also allows his Honour to reach the position, upon which doubt was cast by Young,¹⁰¹ that s 180(2) has some protective work to do in cases where in its absence there would arguably be a contravention of s 180(1). It will be recalled that Young had questioned whether the business judgment rule was really just a form of 'window dressing'.

His Honour then tied together the third and fourth elements of the statutory business judgment rule. He stated:

The director or officer's belief about the best interests of a corporation is to be formed, and its rationality assessed, on the basis of the information obtained through compliance with sub-[s] (c). It is not to be assumed, for the purposes of applying sub-[s] (d), that the director or officer knew everything that he or she ought to have known, but only the things that he or she reasonably had believed to be appropriate to find out.¹⁰²

The above reasoning raises two issues for consideration: the standard applicable to the director's or officer's belief that the judgment is in the best interests of the corporation, and the interaction between ss 180(2)(c) and 180(2)(d).

⁹⁸ Ibid 153–4 [7288]–[7289].

⁹⁹ Ibid 154 [7289].

¹⁰⁰ Ibid.

¹⁰¹ Young, above n 26, 94.

¹⁰² *ASIC v Rich* (2009) 236 FLR 1, 154 [7291].

1 *Rational, Reasonable, or Something In-Between*

The interpretation placed on s 180(2)(d) is not free from doubt, as pointed out by *Ford's Principles of Corporations Law*. This text asks the question of whether, because s 180(2) provides that a director's belief will be a rational one unless the belief is one that no reasonable person in their position would hold, the rationality of the belief 'is to be assessed according to whether it is reasonable'.¹⁰³ The consequence of such a view, the text argues, is that the rule in s 180(2) will have a narrower application than the ALI's business judgment rule.¹⁰⁴ However, the decision in *ASIC v Rich* is quite the opposite.

Consequently, it is necessary to revisit the ALI's position on the rational belief requirement. The ALI was clearly cognisant of the distinction between 'rational' and 'reasonable', recognising that although the words have sometimes been used interchangeably, there is an important distinction to be drawn in their use here:

The phrase 'rationally believes' is intended to permit a significantly wider range of discretion than the term 'reasonable', and to give a director or officer a safe harbour from liability for business judgments that might arguably fall outside the term 'reasonable' but are not so removed from the realm of reason when made that liability should be incurred.¹⁰⁵

However, as explained earlier, the ALI states that the term 'rationally believes' has both an objective and a subjective content. In order to obtain the protection of the business judgment rule, a director or officer must actually believe that the business judgment is in the best interests of the corporation, and that belief must be rational.¹⁰⁶ Further, the ALI's equation of 'rationally believes' with the Delaware 'gross negligence' standard also suggests an objective element, as the negligence test is objective. This rationality standard is much less stringent than a reasonability standard, but nevertheless does involve some, although limited, objective review of the quality of the decision.

The interpretation that *ASIC v Rich* gives to the qualification 'that no reasonable person in their position would hold' at the end of s 180(2) is strained. Essentially, his Honour is saying that where the drafters of the legislation deliberately used the word 'reasonable', they in fact had in mind something quite different to the objective test usually signified by that word.¹⁰⁷ The lack of an objective element may be contrary to the balance between director authority and accountability that the legislature sought, and may even set the bar below that required by the ALI. If 'rational' and 'reasonable' are points on a spectrum, then the last sentence in s 180(2) can be read

¹⁰³ Ford, Austin and Ramsay, above n 18, [8.310].

¹⁰⁴ *Ibid.*

¹⁰⁵ ALI, above n 1, 136. See also *ASIC v Rich* (2009) 236 FLR 1, 153 [7286].

¹⁰⁶ ALI, above n 1, 172.

¹⁰⁷ See, eg, *Corporations Act 2001* (Cth) ss 674(2B), 731.

such that whilst ‘rational’ does not have to mean ‘reasonable’, there is a floor below which directors will lose their entitlement to the benefit of the business judgment rule.

If however, what is required from directors is that they have a rational belief, this might arguably give rise to too low a standard for directors.¹⁰⁸ The United States Third Circuit Court of Appeals has commented that ‘[o]vercoming the presumptions of the business judgment rule on the merits is a near-Herculean task’.¹⁰⁹ If lack of rational belief is akin to gross negligence, then some have asked why the law gives much greater protection to directors, given that ‘ordinary’ negligence is sufficient to create liability for just about anyone else.¹¹⁰ A construction of s 180(2) that arguably sets the bar even lower in Australia, arguably defeats the legislature’s objective of balancing the protection of the authority of directors in the exercise of bona fide business decisions and ensuring they remain accountable to shareholders.

2 *The Interaction Between ss 180(2)(c) and 180(2)(d)*

This article has argued that there should be an objective element to both ss 180(2)(c) and 180(2)(d) in order to strike an appropriate balance between competing policy concerns.

Simply put, a distinction may be drawn between the *process* of arriving at the decision and the *substance* of the decision. If the process by which a decision was made satisfies a reasonability standard — including matters such as the preparations to make a decision, general monitoring, and following up in suspicious circumstances — then the substantive decision itself can be reviewed only under the much looser standard of rationality.¹¹¹ If the director’s enquiry of the subject matter of the decision is reasonable, the court will not interfere. To have both ss 180(2)(c) and 180(2)(d) tested on a subjective standard risks insulating directors from due oversight.

V CONCLUSION

The decision in *ASIC v Rich* should be viewed as resurrecting the business judgment rule as it exists in Australian corporate law, so that the rule is capable of providing a defence in some cases that would otherwise amount to a breach of duty. This

¹⁰⁸ William Lucius Cary and Melvin Aron Eisenberg, *Cases and Materials on Corporations* (Federation Press, 7th ed, 1995) 411:

The rationality standard of review is much easier for a defendant to satisfy than a prudence or reasonability standard. To see how exceptional a rationality standard is, one need only think about the judgments we make in everyday life. It is common to characterise a persons conduct as imprudent or unreasonable, but it is very uncommon to characterise a person’s conduct as irrational.

¹⁰⁹ *In re Tower Air, Inc*, 416 F 3d 229, 238 (3rd Cir, 2005).

¹¹⁰ Gervurtz, above n 49.

¹¹¹ *Ibid* 300–1; Cary and Eisenberg, above n 108, 435.

development is to be applauded. However, the controversy that has long surrounded the implementation of a business judgment rule in Australia is likely to continue through the close and critical consideration of the correctness of the reasoning in *ASIC v Rich*.

This controversy may well be the inevitable result of any judgment applying this rule, given that s 180(2) is a statutory compromise between two competing policy objectives: recognising the director's authority to make decisions without being second-guessed, and ensuring that the director remains accountable. Further, the statutory language is elusive in how it seeks to resolve the competition.

However, although there are inherent difficulties in applying s 180(2), the balance currently struck between a director's authority and accountability is weighted towards the former. This is the result of over-reliance on a subjective state of mind of directors in relation to both the process adopted to inform themselves, and to whether a business judgment is in the best interests of the corporation. Without an objective element in the business judgment rule, the standard of care required of directors may become overly protected and, as a consequence, the objective of accountability is compromised. As a result, Australian policy-makers must revisit the wording of s 180(2) in an attempt to strike an appropriate balance between a director's authority and accountability.

*Mitchell Landrigan**

CAN THE IMPLIED FREEDOM OF POLITICAL DISCOURSE APPLY TO SPEECH BY OR ABOUT RELIGIOUS LEADERS?

ABSTRACT

Religious leaders in Australia sometimes contribute to public debate. Can the implied freedom protect speech by and about religious leaders from a burden (legislative, common law, executive) either current (eg, religious vilification laws) or proposed? In part, the answer to this question depends on whether such expression is recognised as ‘political speech’ under constitutional law. Mason, Toohey and Gaudron JJ’s joint judgment in *Theophanous v Herald & Weekly Times Ltd* (1994) 182 CLR 104 suggests that political speech includes discourse about the political views and public conduct of all persons involved in activities that have become the subject of political debate, a description that could possibly apply to speech by and about religious leaders. By contrast, the High Court’s later unanimous judgment in *Lange v Australian Broadcasting Corporation* (1997) 189 CLR 520 arguably suggests a narrower view of the scope of the implied freedom. It indicates that the implied freedom only protects speech that is relevant to representative government, responsible government or referenda. Is it possible to reconcile *Theophanous* and *Lange*? What do the terms ‘representative government’ and ‘responsible government’ mean? Is it possible to discern a clear view from the High Court judgments about whether speech by or about religious leaders can be political speech?

* Visiting Fellow, Faculty of Law, University of Technology Sydney; PhD Candidate, Law School, University of New South Wales; Solicitor, Supreme Court of New South Wales and High Court of Australia. The author thanks Professor George Williams for his guidance on this area of research as my PhD supervisor, Alison Gurr for her helpful suggestions on an earlier draft, and two anonymous referees for their constructive comments on an earlier draft. This article was in press when the High Court of Australia published its decision in *Unions NSW v NSW* [2013] HCA 58 (18 December 2013). As a result, the article does not consider that decision. Unless otherwise stated, the views expressed in this article are only the author’s and the author is responsible for any errors.

I INTRODUCTION

Some political topics discussed in Australia, such as gay marriage, touch upon religious interests. Australian religious leaders¹ sometimes contribute to public debate.² In view of their contributions to public discourse this article explores whether the speech of and about religious leaders may attract protection from a legislative, common law or executive burden.³ Such burdens *may* include religious tolerance laws or defamation laws if those laws limit constitutionally protected speech by or about religious leaders.⁴

There is a considerable body of academic writing relating to the implied freedom. The literature includes some striking insights into the implied freedom and how it might apply to aspects of religious expression. In this article, however, I chart a mostly unexplored channel of inquiry and ask how the High Court's guidance in relation to protecting political speech might apply to expression by or about religious leaders. Implicit in this investigation is an inquiry into whether speech by or about religious leaders is so qualitatively different to expression about other matters that the implied freedom may not protect such speech. I, therefore, seek to answer the following questions:

- What is political speech?
- Can the implied freedom of political discourse protect speech by and about religious leaders from a burden either current or proposed?⁵

¹ In this article, the phrase 'religious leaders', refers to those in senior roles in organisations whose primary responsibility is 'religious ministry' (by which I mean support and spiritual guidance for a sizeable group of people who share broadly common beliefs about the same god). My interest is, accordingly, in the speech of muftis, deans, bishops and cardinals (typically men) who occasionally (or, depending on their style, possibly more regularly) shift their gaze from spiritual matters to the public square and then contribute to political discussion (however political might be defined) on behalf of those groups, such as might fall within the protective boundaries of the implied freedom. I do not include in this definition those with strong religious affiliations who also happen to be politicians (say, like Reverend Fred Nile), nor do I include those who are primarily involved in charitable work or even semi-political work who may happen to represent religious groups (say, like Reverend Tim Costello).

² See, eg, Cardinal George Pell, Catholic Archbishop of Sydney, 'One Christian Perspective on Climate Change' (Speech delivered at the Annual Global Warming Policy Foundation Lecture, London, 26 October 2011) <<http://www.youtube.com/watch?v=EgsyYXfaIWw>>.

³ Any reference to a burden in this article is a reference to a legislative, executive or common law burden.

⁴ Another possibly relevant burden is blasphemous libel — a prohibition which continues to operate in some Australian jurisdictions. See, eg, s 119 *Criminal Code Act 1924* (Tas).

⁵ This article does not specifically seek to explore aspects of any separation of church and state in Australia nor whether religion belongs in the private sphere.

Answering the first question is a prerequisite to addressing the second question and requires analysis of the underlying legal principles in the relevant High Court cases on the implied freedom of political discourse.

The structure of the article is as follows. In pt II(A), I discuss the views of several writers in relation to defining political speech. In pt II(B), I review whether there are any constitutional limitations to political speech applying to expression about religion, including by reason of s 116 of the Constitution. In pt II(C), I study some of the earlier High Court decisions in relation to characterising political speech and, in pt II(D), I consider the some more recent cases on the implied freedom.

In pt III(A), I analyse in detail the legal principles relating to ‘political discourse’. I note a possible ongoing tension between the kinds of ‘political and governmental matters’ referred to in *Theophanous v Herald & Weekly Times Ltd* (*‘Theophanous’*)⁶ and *Lange v Australian Broadcasting Corporation* (*‘Lange’*)⁷. Then, in pt III(B), I explore what it *means* for the scope of the implied freedom to include speech about representative and responsible government.

In pt IV, I analyse examples of speech by two religious leaders. I review speech by and about the former Australian Mufti, Sheikh Taj El-Din Hamid Hilaly, as well as a response to the Mufti from a Senior Anglican, Dean Phillip Jensen of the Sydney Anglican Diocese. I ask whether the Sheikh’s sermon could be political speech and also consider whether Dean Jensen’s comments about Sheikh Hilaly could fall within the protective boundaries of the implied freedom.⁸ Part V is the conclusion, in which I briefly answer the aforementioned questions.

II WHAT IS POLITICAL SPEECH?

A Commentators’ Views

Dan Meagher argues that political communication consists of any form of expression which ‘may reasonably be relevant to the federal voting choices of its likely audience’.⁹ Meagher bases his approach on his view that ‘a minimalist model of judicially-protected popular sovereignty’ underpins Australia’s implied free speech

⁶ (1994) 182 CLR 104.

⁷ (1997) 189 CLR 520.

⁸ I do not consider in detail whether any *particular* burden — legislative, common law or executive — might operate in relation to the speech of these religious leaders. The purpose of the exercise is to instead consider whether their speech could constitute political speech under constitutional law. So far as I am aware, there was no constraint on either man expressing their views.

⁹ Dan Meagher, ‘What is “Political Communication”? The Rationale and Scope of the Implied Freedom of Political Communication’ (2004) 28 *Melbourne University Law Review* 438, 467.

protections rather than any traditional free speech rationale.¹⁰ From this perspective, speech by or about religious leaders could fall within the protective scope of the implied freedom if it is relevant to the voting choices of its intended audience.

Upon examining the High Court's decision in *Lange*,¹¹ Adrienne Stone advocates redrawing the boundaries of the scope of the implied freedom in order to facilitate the *proper* functioning of the institutions the implied freedom seeks to protect.¹² In doing so, Stone proposes four kinds of speech that may be regarded as 'political communication' while acknowledging that these proposed categories of free speech may tend to dramatically expand the scope of protected speech.¹³

The first category Stone proposes is explicitly political communication — ie, speech substantively about the government.¹⁴ The second kind of political communication Stone posits is speech about issues that could *become* matters of federal law or policy or, which in some other way, could be the subject of federal governmental action.¹⁵ This category anticipates the possibility for change in a current political agenda and could include discourse that *becomes* relevant to governmental decision-making.¹⁶ A third type of political communication suggested by Stone is communications about matters which may not themselves be the subject of law or government action but which may nonetheless influence voters' attitudes toward the government.¹⁷ Finally, Stone suggests as a fourth class of political communication speech which is 'relevant to democratic government because of the qualities it develops in the citizenry'.¹⁸ Stone describes this kind of political speech as that which 'develop[s] among voters the capacities or qualities necessary to make a 'true choice' in a federal election'.¹⁹

In studying the scope of the implied freedom, Nicholas Aroney²⁰ traces the influential judgments of McHugh J in the High Court.²¹ Aroney observes that his Honour

¹⁰ Ibid 438.

¹¹ (1997) 189 CLR 520.

¹² Adrienne Stone 'Rights, Personal Rights and Freedoms: The Nature of the Freedom of Political Communication' (2001) 25 *Melbourne University Law Review* 374, 383.

¹³ Ibid 383–4.

¹⁴ Ibid 384.

¹⁵ Ibid 385.

¹⁶ Ibid.

¹⁷ Ibid 386.

¹⁸ Ibid 387.

¹⁹ Ibid.

²⁰ See, eg, Nicholas Aroney 'Justice McHugh, Representative Government and the Elimination of Balancing' (2006) 28 *Sydney Law Review* 505; Nicholas Aroney 'The Structure of Constitutional Revolutions: Are the *Lange*, *Levy* and *Kruger* Cases a Return to Normal Science?' (1998) 21 *University of New South Wales Law Journal* 645.

²¹ See, eg, *Australian Capital Television v Commonwealth* (1992) 177 CLR 106; *Theophanous v Herald & Weekly Times Ltd* (1994) 182 CLR 104; *McGinty v Western Australia* (1996) 134 ALR 289.

focused on the systems of representative government and responsible government as the twin foundations for delineating the scope of the implied freedom. McHugh J maintained that these systems are not free-ranging principles but are supported by the text and structure of the Constitution.²² Further, McHugh J distinguished representative government — generally understood to mean that people in free elections choose representatives who govern on their behalf — from *representative democracy* — a wider concept, often taken to be descriptive of a society in which there is an equality of rights and privileges.²³

A paper by David Hume about the rationale for the implied freedom suggests that it protects the receipt of relevant information by those responsible for making constitutionally-prescribed decisions.²⁴ Hume observes that responsible government, representative government and referenda each involve the making of political choices. He contends that the efficacy of these systems requires that government not restrict the choosers' opportunity to access relevant information concerning the choices contemplated by the *Constitution* and that *this* makes the implied freedom necessary.²⁵ On this view, the implied freedom is a tool for facilitating the making of informed choices which support the systems of responsible government, representative government and referenda.²⁶

Two broad themes emerge from this brief analysis of some of the literature.

First, as Meagher suggests, political speech may consist of speech that is relevant to an audience's federal voting choices. This is a useful touchstone for describing the protective scope of the implied freedom. Yet, the High Court under McHugh J's influence has taken the view that speech which is relevant to representative government *and* responsible government²⁷ may fall within the constitutional freedom. A conception of the implied freedom that is grounded only in speech relevant to the audience's *voting choices* may not adequately reflect the scope of the freedom as it has been described by the High Court.

Secondly, speech by or about religious leaders may come within the protective scope of the constitutional freedom if the (burdened) discourse is relevant to the Commonwealth systems of representative government and responsible government. Stone, as noted, suggests that the High Court in *Lange* narrowed the protective scope of the implied freedom. While this may be so, I propose that there may be fewer differences

²² Aroney, 'Representative Government and Elimination of Balancing', above n 20, 514.

²³ *Ibid.*

²⁴ David Hume, 'Wotton v Queensland — "Islands of Power" and Political Speech on Palm Island' (Paper presented at the 2012 Constitutional Law Conference, Faculty of Law Research Series 2012, University of New South Wales, 17 February 2012).

²⁵ *Ibid.* 8.

²⁶ *Ibid.* 9.

²⁷ The constitutional freedom may also include speech relevant to referenda. Any further reference to representative and responsible government includes a reference to referenda.

between *Theophanous* and *Lange* than might at first seem apparent, particularly if the courts undertake a generous temporal perspective of representative and responsible government. Further, both decisions have the potential to recognise speech by and about religious leaders as political speech.

B Are there Constitutional Limitations to Speech about Religion being Political Speech?

Does the *Commonwealth Constitution* prevent the implied freedom applying to speech about religion?²⁸ Aroney explores whether there may be constitutional barriers to the implied freedom protecting speech about religion.²⁹ He specifically examines whether s 116 of the *Constitution* might imply a barrier to the implied freedom protecting religious expression.³⁰ Aroney's line of inquiry is this: if s 116 of the *Constitution* prevents the Commonwealth from making religiously based laws (including laws for establishment of religion or for prohibiting the free exercise of religion), then this may mean religious speech cannot form a constitutionally legitimate part of political communication. This in turn may suggest that, *by definition*, the implied freedom cannot protect speech about religion.³¹

After carefully reviewing the legal authorities and academic literature, Aroney concludes that there are no a priori reasons why speech about religious matters cannot simultaneously be characterised as political communication for the purposes of the implied freedom.³² I agree with Aroney's conclusion. I elaborate upon his reasoning and consider other possible a priori limitations on the implied freedom applying to speech about religion — but, as will be seen, I then discount these propositions.

To provide further context to this inquiry — s 116 purports to constrain the Commonwealth government's legislative powers. It provides that the Commonwealth (though not the states or territories) cannot make laws for proscribed purposes or require a religious test as a qualification for office.³³ The 'free exercise' limb of

²⁸ If the *Constitution* poses no barrier to the implied freedom applying to expression about religion, then it would seem to follow that there is no a priori obstacle to the implied freedom applying to speech by or about religious leaders.

²⁹ Nicholas Aroney, 'The Constitutional (In)Validity of Religious Vilification Laws: Implications for their Interpretation' (2006) 34 *Federal Law Review* 288, 292.

³⁰ Section 116 provides that the Commonwealth shall not make any law for establishing any religion, or for imposing any religious observance, or for prohibiting the free exercise of any religion; and no religious test shall be required as a qualification for any office or public trust under the Commonwealth. To date, there has been no successful High Court action brought under s 116.

³¹ Aroney, 'Constitutional (In)Validity of Religious Vilification Laws', above n 29, 292.

³² Cf *Harkianakis v Skalkos* (1997) 47 NSWLR 302, 307 (Dunford J) (also referred to in *ibid*).

³³ The provision does not purport to apply to rules or standing orders or exercises of executive power.

s 116³⁴ specifically prohibits the Commonwealth making *laws* for *prohibiting* the free exercise of religion. It does not apply to exercises of executive power and instead governs the making of *laws* which seek to effect *prohibitions* on religious expression.³⁵ The High Court has not given a broad interpretation to the free exercise provision or s 116 more generally.³⁶

Given these apparently minimal constraints on Commonwealth legislative power, it is hard to see why s 116 should lead to any excision of religious speech from the protective scope of the implied freedom. Further, as Aroney notes, s 116 could be amended by referenda and the prohibitions in s 116 could not circumscribe the scope of the implied freedom of political communication in relation to potential changes to s 116.³⁷ Accordingly, there is no strong argument as to why any ‘rights’ conferred by s 116 may limit the operation of the implied freedom with respect to speech about religion.

A further argument (which I also discount) may be made: that s 116 limits the reach of the implied freedom in relation to speech about religion when state or territory laws are involved. Section 116 does not prohibit the states or territories from making laws for religion. It might therefore be argued that the states and territories should be able to administer laws relating to religious matters without the implied freedom encroaching upon their constitutional powers.³⁸ It may even be said that if the implied freedom were to *protect* the kind of religious speech that the states are authorised to *prohibit* (say, vilifying speech that is regulated under state-based religious tolerance laws), then, at that point, the implied freedom has reached too far, constitutionally speaking.

Even if one accepts that the states and territories *can* make laws about religious affairs (by reason of s 116 or otherwise), there are significant limitations with this ‘overreach’ logic. Specifically, this reasoning has two weaknesses. First, even *if* a state government

³⁴ This would seem to be the subclause of s 116 that plaintiffs would most likely invoke to protect religious expression. A plaintiff may also, or alternatively, plead the ‘no establishment’ (or even the ‘no imposing religious observance’) limb of s 116 to support a constitutional ‘religious expression’ claim.

³⁵ Aroney, ‘Constitutional (In)Validity of Religious Vilification Laws’, above n 29, 300.

³⁶ See, eg, *Adelaide Company of Jehovah’s Witnesses Inc v Commonwealth* (1943) 67 CLR 116; *Krygger v Williams* (1912) 15 CLR 366. The provision arguably would not apply to laws (or executive acts) which merely seek to stifle or suppress (cf *prohibit*) the free expression of religion.

³⁷ Aroney, ‘Constitutional (In)Validity of Religious Vilification Laws’, above n 29, 301.

³⁸ In *APLA Ltd v Legal Services Commissioner (NSW)* (2005) 224 CLR 322, Callinan J at 478 expressed concern that the definition of political speech could, if not ‘bounded’, include religious speech and might thereby conflict with the states’ powers to make laws for religion under s 116 of the *Constitution*. Arguably s 116 does not confer any powers over religion upon the states or territories (although it may do by implication). There were suggestions during the Constitutional Convention debates that some framers likely *did* intend to preserve to the states the power to enact Sunday observance laws: author PhD research. The states also may have plenary or more limited powers in relation to religion under their own ‘constitutions’.

has constitutional power over a topic, people may have an interest in discussing *how* that government exercises its powers with respect to that topic. An assumed limitation on constitutional free speech in relation to either state laws or state government actions may lead to voters having no constitutional free speech protection if the states happen to prosecute religious agendas.³⁹ The implied freedom would become a hollow constitutional safeguard if it could not, *by definition*, protect speech about state governments that introduce laws affecting or relating to religious interests. A second reason is a pragmatic one relating to the likelihood that the implied freedom would regularly interfere with states' powers. As will be seen more clearly after reviewing some of the High Court cases — to *be* political speech, the expression must in some way relate to the *Commonwealth* systems of representative and responsible government. Speech about religion in relation to a state law which has *no* implications for the Commonwealth would likely not qualify for the constitutional protection.

There is a final argument that is similar (but subtly different) to that explored by Aroney which could be an alternative reason for the implied freedom not protecting speech about religion. It might be observed that the Commonwealth has no specifically enunciated constitutional power over religion (notwithstanding that it may be able to pass laws with respect to religion if they are also with respect to a matter or purpose within power, say in areas such as naturalisation and immigration/emigration⁴⁰).⁴¹ It might *then* be posited as part of this argument that political discourse — speech about the government — *ought not to* include speech about religion. Accordingly, as a kind of counterpart to Aroney's argument, it may be contended that the Commonwealth's *lack* of legislative powers over religion is *itself* a reason to exclude speech about religion from the protective scope of the implied freedom.

In my view, however, it could not be argued — at least not with conviction — that an absence of express Commonwealth power over religion ought to lead to any excision of speech about religion from the protective scope of the implied freedom. To begin with, the lack of legislative powers over religion is no guarantee that a government will eschew religious objectives.⁴² Excluding speech about religion

³⁹ Such an approach would arguably also be at odds with the High Court's acceptance that the implied freedom *can* protect speech that is affected by state (or territory) laws. See, eg, *Coleman v Power* (2004) 220 CLR 1.

⁴⁰ See *Official Report of the National Australasian Convention Debates*, Melbourne, 2 March 1898, 1771–72 (Edmund Barton).

⁴¹ While the Preamble was not intended to be an operative provision (or to confer any direct power on the Commonwealth), the Preamble recognises 'god' by acknowledging an Almighty God.

⁴² The federal Labor government's funding of school chaplains is arguably evidence of the pursuit of quasi religious objectives, particularly in light of then Prime Minister Gillard agreeing with the suggestion of the head of the Australian Christian Lobby (Jim Wallace), in an interview on 6 August 2010, that there was little doubt that the chaplaincy program would continue after 2011 even though the program's continuation at the time was yet to be reviewed. See Interview with Jim Wallace, Christian Lobby, (Make It Count 2010, 6 August 2010) <<http://www.acl.org.au/make-it-count/>>; *Williams v Commonwealth* (2012) 288 ALR 410.

from the scope of the implied freedom would potentially mean there would be *no* protection for speech about *how* the government exercises any quasi-religious powers. Such a restriction on the limitations of political speech would be antithetical to any notion that the implied freedom can promote the government's accountability to the electorate by allowing voters to share and exchange political ideas, including — perhaps most critically in this context — if a government seeks to implement religiously motivated policies.

It follows from what I have said in this part that there are no 'in principle' reasons as to why the implied freedom cannot apply to speech about religion or to expression by or about religious leaders. As will also be seen, no legal limitations about the implied freedom applying to speech about religion or to speech by or about religious leaders appear to emerge from the High Court cases on the implied freedom.

C The High Court's Early Views on Political Speech

In *Nationwide News Pty Ltd v Wills* ('*Nationwide News*')⁴³ and *Australian Capital Television v Commonwealth* ('*ACTV*'),⁴⁴ the High Court recognised an implied freedom in relation to speech about governmental and political matters. In 1994, in *Theophanous*⁴⁵, the High Court offered a more detailed insight into what political speech means. Writing a joint judgment, Mason CJ, Toohey and Gaudron JJ (in the majority, with Deane J) said political speech includes:

discussion of the conduct, policies or fitness for office of government, political parties, public bodies, public officers and those seeking public office. The concept also includes discussion of the political views and public conduct of persons who are engaged in activities that have become the subject of political debate, eg, trade union leaders, Aboriginal political leaders, political and economic commentators. Indeed, in our view, the concept is not exhausted by political publications and addresses which are calculated to influence choices.⁴⁶

Left unaltered by later High Court judgments, this description of political discussion would indicate that the implied freedom applies to speech about the political views and public conduct of all persons involved in activities that have become the subject of political debate, possibly including religious leaders. It is a persuasive judgment, being jointly written by three judges in a majority High Court decision and has never been expressly overruled.

Even though the application of the implied freedom to *state* political matters was not directly in issue in *Theophanous* (it was a defamation action brought against a newspaper by a Commonwealth politician), Mason CJ, Toohey and Gaudron JJ indicated that the implied freedom could not be confined to matters about the

⁴³ (1992) 177 CLR 1.

⁴⁴ (1992) 177 CLR 106.

⁴⁵ (1994) 182 CLR 104.

⁴⁶ *Ibid* 123.

Commonwealth government, holding that it would be ‘unrealistic’ to attempt to confine the freedom to matters relating to the Commonwealth government.⁴⁷

The question of whether the implied freedom applied to state political affairs was considered by the High Court in another defamation case — *Stephens v West Australian Newspapers Limited* (‘*Stephens*’)⁴⁸ — decided shortly after *Theophanous*. *Stephens* concerned six members of the Legislative Council of Western Australia who took an overseas trip as members of a standing committee on state government agencies. The *West Australian* newspaper implied that the six state politicians wasted public money on an overseas trip in order to investigate matters which, it was claimed, could have been conducted in Western Australia.⁴⁹

Consistent with the view they expressed in *Theophanous*, Mason CJ, Toohey and Gaudron JJ jointly held in *Stephens* that the implied freedom also applied to state political affairs.⁵⁰ Deane J, the fourth member of the High Court’s majority (as he had been in *Theophanous*), noted that state laws may burden freedom of communication on Commonwealth or state matters.⁵¹ Brennan J, dissenting, could not see any obvious connection between the political affairs of the Commonwealth and the Western Australian politicians.⁵² His Honour held that the publication of the material was unaffected by the implied freedom.⁵³ In doing so, Brennan J hinted that the implied freedom applied to Commonwealth political matters and state political affairs only insofar as the latter related to Commonwealth powers.

After the High Court’s majority decisions in *Theophanous* and *Stephens*, it may have seemed that the implied freedom could apply to a very wide range of political matters. Yet, in some of the High Court cases that followed — in large part due to the influence of McHugh J⁵⁴ — the Court began to more closely link the application of the implied freedom to the terms of the *Commonwealth Constitution*. The High Court also began to assess more rigorously whether the protective scope of the implied freedom conformed to the systems of representative and responsible government as these principles operated under the *Commonwealth Constitution*.

In 1997 in the High Court’s landmark decision in *Lange*, seven judges jointly held that *Theophanous* and *Stephens* precluded any further unqualified operation of

⁴⁷ Ibid 122.

⁴⁸ (1994) 124 ALR 80.

⁴⁹ Ibid 92 (Brennan J).

⁵⁰ Ibid 88. In their joint judgment in *Stephens*, Mason CJ, Toohey and Gaudron JJ accepted that a constitutional defence could be available to the defendants under the *Commonwealth Constitution* and/or under the *Constitution of Western Australia*.

⁵¹ Ibid 108.

⁵² Ibid 91.

⁵³ Ibid.

⁵⁴ See, eg, *McGinty v Western Australia* (1996) 134 ALR 289, 348 (McHugh J); Aroney, ‘Representative Government and Elimination of Balancing’, above n 20.

the law of defamation with respect to political discussion and, thereafter, the High Court required recognition of the implied freedom in defamation suits concerning publications about political affairs.⁵⁵ The Court described the test for identifying whether a law is consistent with the implied freedom in the following terms:

[w]hen a law of a State or federal Parliament or a Territory legislature is alleged to infringe the requirement of freedom of communication imposed by ss 7, 24, 64 or 128 of the *Constitution*, two questions must be answered before the validity of the law can be determined. *First, does the law effectively burden freedom of communication about government or political matters either in its terms, operation or effect?* Second, if the law effectively burdens that freedom, is the law reasonably appropriate and adapted to serve a legitimate end the fulfilment of which is compatible with the maintenance of the constitutionally prescribed system of representative and responsible government and the procedure prescribed by s 128 for submitting a proposed amendment of the *Constitution* to the informed decision of the people ... If the first question is answered 'yes' and the second is answered 'no', the law is invalid.⁵⁶

The italicised question in the above passage — the first step in the *Lange* test — asks whether a 'law effectively burden[s] freedom of communication about government or political matters either in its terms, operation or effect?'⁵⁷ The answer to this question depends on what is meant by 'government or political matters'.

In addressing the issue of *what is meant by* government or political matters, the High Court held that the protective scope of the implied freedom applies to communications that are relevant to people making voting choices (ie, representative government). Yet, the judges also found that the implied freedom extends to communications about *responsible* government, stating that:

[the implied freedom] necessarily implies a limitation on legislative and executive power to deny the electors and their representatives information concerning the executive branch of the government throughout the life of a federal Parliament. Moreover, the conduct of the executive branch is not confined to Ministers and the public service. It includes the affairs of statutory authorities and public utilities which are obliged to report to the legislature or to a Minister who is responsible to the legislature.⁵⁸

Responsible government therefore comprises speech relating to the actions of the Commonwealth government and its ministers, including through departments, statutory authorities and public utilities over which ministers have responsibility.

⁵⁵ *Lange* (1997) 189 CLR 555.

⁵⁶ *Ibid* 567 (emphasis added).

⁵⁷ *Ibid*.

⁵⁸ *Ibid* 559.

Lange thus emerged as authority for the proposition that communication about government or political matters includes speech that facilitates informed voting choices and discourse which is relevant to the effectiveness of the government, including through its statutory authorities and public utilities. Yet, *Lange* did not overrule *Theophanous*. The cases arguably stand for different propositions, with the earlier case possibly allowing greater latitude to recognise as political speech the contributions of those (eg, church leaders) who enter into, or contribute towards, public debate. Whether the cases are substantively dissimilar on closer analysis is a question I consider further in pt III.

D *The High Court's Later Views on Political Speech*

In later implied freedom cases, the High Court ruled, for example, that speech was not limited to the written or spoken word and that it could include other kinds of expressive acts,⁵⁹ even possibly public acts of prayer.⁶⁰ In *APLA Ltd v Legal Services Commissioner (NSW)*⁶¹ (*'APLA'*), the High Court heard a challenge to restrictions on the advertising of legal services by regulations introduced under NSW legislation, namely the *Legal Profession Act 1987* (NSW). The applicants in *APLA* argued that cl 139 of the *Legal Profession Regulation 2002* (NSW) (*'Legal Professional Regulation'*)⁶² infringed the implied freedom. By majority, however, the High Court upheld the validity of the clause.

The five majority judges — Gleeson CJ, Gummow, Hayne, Callinan and Heydon JJ — considered that *Lange* did not support protecting the commercial matters that were the subject of the *Legal Profession Regulation*. In a joint judgment, Gleeson CJ and Heydon J wrote that they did not accept that the *Legal Profession Regulation* infringed the implied freedom merely because the regulation might restrict the mentioning of some government or political issue, or prohibit the naming of a particular politician.⁶³ Their Honours held that the implied freedom does not protect communications that are an essentially commercial activity.⁶⁴ By similar reasoning, it might be said that the implied freedom does not protect communications that are essentially or predominantly *religious* in character.

McHugh J in *APLA*, though dissenting, observed that the *Lange* freedom arises from the need to promote and protect representative and responsible government.⁶⁵

⁵⁹ See *Kruger v Commonwealth* (1997) 190 CLR 1; *Coleman v Power* (2004) 220 CLR 1.

⁶⁰ See *Levy v Victoria* (1997) 189 CLR 579, 631 (Kirby J) (*'Levy'*). Toohey and Gummow JJ also recognised that the constitutional freedom is not limited to verbal activity: at 613.

⁶¹ (2005) 224 CLR 322.

⁶² The *Legal Profession Regulation 2002* (NSW) has since been repealed and replaced with the *Legal Profession Regulation 2005* (NSW).

⁶³ *APLA* (2005) 224 CLR 322, 351.

⁶⁴ *Ibid.*

⁶⁵ *Ibid* 362.

Referring to the scope of the implied freedom, his Honour described these twin governmental systems as necessitating ‘some level of communicative freedom in Australian society about matters relevant to executive responsibility and an informed electoral choice’.⁶⁶ Accordingly, McHugh J confirmed that the implied freedom protects: first, discussion about the actions of government and, secondly, communication that facilitates the making of informed voting choices.⁶⁷

In 2011, in *Hogan v Hinch*⁶⁸, the High Court considered the application of the implied freedom to Victorian state-based suppression orders. The Melbourne broadcaster, Derryn Hinch, was charged with contravening suppression orders made by the County Court of Melbourne. The High Court unanimously dismissed Hinch’s claim. French CJ described the implied freedom as protecting speech about matters ‘potentially within the purview of government’, which might, with flavours of the joint judgment in *Theophanous*, imply a more generous view of the scope of the implied freedom than might be otherwise suggested by the *Lange* decision.⁶⁹ Referring to the ‘significant interaction between the different levels of government in Australia’⁷⁰, French CJ quoted the following passage in *Lange* when rejecting counsel’s contention that the implied freedom only applied to politics at the Commonwealth level:

this Court should now declare that each member of the Australian community has an interest in disseminating and receiving information, opinions and arguments concerning government and political matters that affect the people of Australia.⁷¹

French CJ also cited *Lange* to support his view that the implied freedom applied to other levels of government as well as the Commonwealth. His Honour considered this conclusion to be inevitable in view of the existence of national political parties operating at all levels of government, the financial dependence of state, territory and local governments on federal funding and policies, and the increasing integration of social, economic and political matters in Australia.⁷²

Arguably, the above quoted *Lange* passage does not describe the jurisdictional scope of the implied freedom per se. The extract describes the extended political category of qualified privilege that the High Court developed in *Lange* in relation to the law of

⁶⁶ Ibid.

⁶⁷ In *Aid/Watch Incorporated v Commissioner of Taxation* (2010) 272 ALR 417, 429 (*Aid/Watch*), French CJ, Gummow, Hayne, Crennan and Bell J endorsed and reinforced *Lange*’s description of the scope of the implied freedom. (The case was about the tax status of a charity with political purposes).

⁶⁸ (2011) 275 ALR 408.

⁶⁹ Ibid 426.

⁷⁰ Ibid 425.

⁷¹ Ibid 426.

⁷² Ibid.

defamation. On the other hand, if the High Court developed the concept of qualified privilege to ensure compliance *with the Constitution*, then it is perhaps harder to see why the constitutional protection should not extend to the same extent as the qualified privilege.⁷³ French CJ nevertheless (again citing *Lange*) described the implied freedom as protecting discourse that ‘bear[s] on the choice that the people have to make in federal elections or in voting to amend the *Constitution*, and on their evaluation of the performance of federal Ministers and their departments’.⁷⁴ His Honour thus affirmed that the foundation of the implied freedom is in the Commonwealth systems of representative government and responsible government, and it may protect speech that is relevant to people’s *choices* with respect to those systems.⁷⁵

In *Wotton v Queensland* (‘*Wotton*’),⁷⁶ the plaintiff (Wotton), an Aboriginal and political protester, participated in a riot on Palm Island in 2004 following the death in custody of another Aboriginal man. Wotton challenged the provisions of the *Corrective Services Act 2006* (Qld) (‘*Corrective Services Act*’) (which imposed limits on his speech as a parole condition) on the basis that the provisions impermissibly burdened freedom of political communication. The High Court unanimously dismissed Wotton’s application, reasoning that the relevant provisions of the state legislation complied with the constitutional limitation upon the legislative power of the state.

In a joint judgment, French CJ, Gummow, Hayne, Crennan and Bell JJ (with whom Kiefel J generally concurred), described their task in *Wotton* as analysing the nature of any restraint the implied freedom imposed upon the legislative power of the Queensland legislature and examining whether the *Corrective Services Act* conformed to that restraint.⁷⁷ Their Honours differentiated this exercise from that of assessing an alleged ultra vires exercise of discretionary power under legislation — a claim which the judges held lies in administrative law review of that decision.⁷⁸

Concerning the first limb of *Lange* (ie, whether the expression was political speech), the joint judgment held that the affected speech *was* relevant to the Commonwealth tier of government. Their Honours found that public discussion of matters relating to Aboriginal and Indigenous affairs, including perceived or alleged injustices, involves communication at a national level, rather than a purely state

⁷³ Assuming that is, that the suppressed speech is sufficiently relevant to the systems of representative and responsible government.

⁷⁴ *Hogan v Hinch* (2011) 243 CLR 506, 543–4 [49].

⁷⁵ The other six judges in *Hogan v Hinch* — Gummow, Hayne, Heydon, Crennan, Kiefel and Bell JJ — published a joint judgment concurring with French CJ in the result. While their Honours were prepared to accept that the Victorian legislation *may* offend the implied freedom, in applying the second limb of *Lange*, their Honours held that the legislation did not display any ‘direct’ burden upon that communication, as opposed to an incidental one: at 555–6 [95]. The High Court would return to this theme in *Wotton v Queensland* (2012) 285 ALR 1.

⁷⁶ (2012) 285 ALR 1.

⁷⁷ *Ibid* 8.

⁷⁸ *Ibid*.

level.⁷⁹ The five judges also noted that law enforcement and policing depends on cooperation between federal, state and territory police forces and that the interaction between those services and Aboriginal people is a matter of national rather than purely local political concern.⁸⁰ Their Honours thus emphasised the connections between the state laws and the *federal* system of government.

In 2013, the High Court decided *Attorney-General (SA) v Corporation of the City of Adelaide and Others* ('*A-G v Adelaide*').⁸¹ Caleb and Samuel Corneloup, members of Street Church (Caleb was President), preached the Christian gospel in Rundle Mall, a central retail area in Adelaide. A City of Adelaide by-law prohibited, inter alia, any person from 'preaching', 'canvassing' or 'haranguing' on any street or thoroughfare in the City of Adelaide without a permit from the Council.⁸² The Corneloup brothers challenged the by-law in the District Court of South Australia on the bases that the relevant provisions of the by-law were: first, outside the by-law making power of the primary legislation and, secondly, infringed their implied freedom of political discourse. After the brothers' initial legal success in the District Court⁸³ and then after they successfully defended an appeal by the City of Adelaide in the Full Court of the South Australian Supreme Court,⁸⁴ a majority of the High Court⁸⁵ heard and allowed an appeal by the Attorney General for South Australia. The High Court reversed the decision of the Full Court and upheld the validity of the City of Adelaide by-law.

⁷⁹ Ibid 9.

⁸⁰ Ibid. Examining the second limb of *Lange*, the five judges referred to the distinction the High Court made in *Hogan v Hinch*, namely, between a law which only incidentally restricts political communication and a law which prohibits or regulates communications which are political or are a necessary ingredient of political communication. Stating that the Queensland laws in this instance were in the former category, the five judges held that the burden upon communication is more readily seen to satisfy the second limb of *Lange* if the law is of the former rather than the latter description.

⁸¹ (2013) 295 ALR 197. The High Court consisted of six judges.

⁸² Ibid 199.

⁸³ Finding in the brothers' favour, his Honour Judge Stretton in the District Court held that the by-law was not authorised under the relevant primary legislation. Judge Stretton did not consider it necessary to review whether the by-law also infringed the implied freedom. See *Corneloup v Adelaide City Council* [2010] SADC 144 (25 November 2010).

⁸⁴ On appeal, the Full Court of the South Australian Supreme Court held (unanimously) that the by-law was a valid exercise of the power under the primary legislation but that the by-law *did* infringe the implied freedom. It thus also found in the brothers' favour. See *Corporation of The City of Adelaide v Corneloup* (2011) 110 SASR 334.

⁸⁵ Before the High Court, the Corneloups argued that the by-law was ultra vires the primary legislation, that proper procedural steps were not followed in the making of the by-law and that the by-law infringed the implied freedom. French CJ, Hayne, Crennan, Kiefel and Bell JJ (the latter agreeing with Hayne J) held that the by-law was authorised under the primary legislation and that it did not infringe the implied freedom. Heydon J, dissenting, ruled that the by-law was not authorised under the primary legislation. I do not review in detail the aspects of the decision relating to the claim that the by-law was not a valid exercise of power under the primary legislation.

Before the High Court, the Solicitor General for South Australia conceded that the provisions of the City of Adelaide by-law could burden communications about political communications.⁸⁶ The concession, potentially limited the opportunity for the Court to determine whether a law which restricts preaching has *such* a slight or inconsequential effect on political speech as to not burden it.⁸⁷ This point is borne out by examining the speech ostensibly burdened in this case. Caleb Corneloup's affidavit stated that his speech related to federal affairs (eg, gay marriage, internet filtering, teenage binge drinking and abortion).⁸⁸ These topics plainly seem to be related to federal (or federal and state) politics. Caleb, however, also preached about 'the unscientific nature of evolution being taught in schools etc', and the 'supremacy of god over man'; his brother Samuel Corneloup described himself as an 'expositor of the gospel'.⁸⁹ Arguably, the brothers' speech was only peripherally connected with politics because — however well expressed — it was a message of Christian salvation. The High Court did not explain how preaching a message about eternal life in Jesus Christ might relate to representative or responsible government in the Commonwealth Parliament.

The High Court published another implied freedom case — *Monis*⁹⁰ — on the same day as *A-G v Adelaide*. In *Monis*, the appellant, Man Haron Monis, was alleged to have sent letters and CDs to the families and relatives of those Australian soldiers

⁸⁶ See *A-G v Adelaide* (2013) 295 ALR 197, 221 (French CJ). French CJ noted that preaching, canvassing, haranguing and the distribution of literature are all activities which may be undertaken in order to communicate to members of the public matters and which may be directly or indirectly relevant to politics or government at the Commonwealth level: at 222. Hayne J was prepared to assume that because the by-law prohibited certain conduct on a road without permission that it effectively burdened political communication: at 236. Crennan and Kiefel JJ, in a joint judgment, noted that it was not disputed that the requirement to obtain a permit effected a burden on some communications of a political kind sought to be made in the process of preaching or canvassing: at 254. Despite the Solicitor General's concession in relation to the first limb of *Lange*, French CJ, Hayne, Crennan and Kiefel JJ (Heydon J not deciding on this point and Bell J concurring with Hayne J) held that the subject provisions of the by-law were consistent with the second limb of *Lange*.

⁸⁷ See *Monis v The Queen* (2013) 295 ALR 259 ('*Monis*'). In *Monis*, published on the same day as *A-G v Adelaide*, Crennan, Kiefel and Bell JJ noted that an 'effect upon political communication which is so slight as to be inconsequential may not require an affirmative answer to the first limb enquiry': at 344 [343].

⁸⁸ See affidavit cited in the District Court proceedings in *Corneloup v Adelaide City Council* [2010] SADC 144 (25 November 2010) [8]. Caleb also claimed to be 'working closely with and on behalf of Joseph Stephen ... a candidate for "The Christian Democratic Party"': at *ibid*.

⁸⁹ See *Corneloup v Adelaide City Council* [2010] SADC 144 (25 November 2010) [8], [10]. Samuel Corneloup refused to apply for a preaching permit because he said it was against his religious beliefs to ask for permission to do so: at [10]. Street Church's message was about the authority of the Christian Bible: see, eg, Vassil Malandris from ABC Television, Interview with Adelaide Street Preachers Caleb and Samuel Corneloup (Television Interview, 19 March 2013) <http://www.youtube.com/watch?v=J6-R_YIcP6Y>.

⁹⁰ (2013) 295 ALR 259.

who were killed during active duty in Afghanistan. The letters typically began by expressing condolences to the recipients and then described each deceased soldier in increasingly critical terms as, being amongst other things, murderers. Copies of some of the correspondence were sent to senior political leaders.

Monis was charged with offences under s 471.12 of the *Criminal Code 1995* (Cth).⁹¹ The second appellant, Ms Amirah Droudis, was alleged to have aided and abetted Monis in relation to some of the same offences.⁹² After unsuccessfully trying to have the allegations against them quashed in both the NSW District Court and the NSW Court of Criminal Appeal on the basis that s 471.12 of the *Criminal Code* infringed the implied freedom, Monis and Droudis lodged a further appeal with the High Court arguing that the provision infringed the implied freedom.

The High Court (consisting of six judges as it had done in *A-G v Adelaide*) held unanimously that s 471.12 of the *Criminal Code* burdened political communication.⁹³ Yet, the six judges were evenly divided on the question of whether the law satisfied the test in the second limb of *Lange*. In three separate judgments, French CJ, Hayne and Heydon JJ decided that the law was inconsistent with the second limb of *Lange* (their Honours therefore declared s 471.12 invalid and would have granted the appeal). In a joint judgment, however, Crennan, Keifel and Bell JJ held that the law was consistent with the second limb of *Lange* and their Honours rejected the appeal.⁹⁴

All the judges in *Monis* concluded that s 471.12 of the *Criminal Code* — which regulated offensive uses of postal and similar services — burdened political

⁹¹ This provision prevents the use of a postal or similar service in a way that reasonable persons would regard as being, in all the circumstances, menacing, harassing or offensive. Twelve of the charges against Monis related to ‘offensive’ use of the postal service, and one related to using the postal service to harass. The appellants did not pursue an appeal on constitutional grounds in relation to the alleged harassment offence. The appeal from the District Court (and the later appeal) was concerned with the constitutionality of s 471.12 insofar as it concerned ‘offensive’ uses of a postal service.

⁹² At the time of the High Court hearing, the appellants were yet to be tried on the substantive offences.

⁹³ See *Monis* (2013) 295 ALR 259, 280–2 (French CJ), 287 (Hayne J), 318 (Heydon J), 344 (Crennan, Keifel and Bell JJ).

⁹⁴ The appeal therefore failed. Section 23(2)(a) of the *Judiciary Act 1903* (Cth) provides that where the High Court is evenly divided with respect to its opinion on a lower court’s decision, the decision of the lower court is affirmed. Had the appellants sought to challenge the validity of the legislation in the original jurisdiction of the High Court, the same decision would have resulted in their favour, since that decision would have been governed by s 23(2)(b) of the *Judiciary Act 1903* (Cth) which provides that if judicial opinion is divided in a case brought in the original jurisdiction of the High Court, then the Chief Justice’s judgment will prevail. The appellants later pleaded guilty to the offences. Monis was sentenced to 300 hours of community service and received a two year good behaviour bond. Droudis received a two year good behaviour bond for aiding and abetting Monis in relation to one count.

speech.⁹⁵ Chief Justice French noted that some political communications may cause disgust, hatred or outrage.⁹⁶ Hayne J similarly reasoned that constitutionally protected speech can be offensive. His Honour held that invective as well as abuse can be part of political discourse.⁹⁷

On the question of what constitutes a burden on political speech, the Commonwealth Director of Public Prosecutions (as well as the Commonwealth and the States of Queensland, South Australia and Victoria) submitted to the High Court that some restraints on speech apply to such a narrow category of political communication that the burden *cannot* be inconsistent with the implied freedom because it would not effectively burden political communication.⁹⁸ Dismissing this argument, Hayne J (supported by other members of the High Court) held that ‘effectively burden’ means no more than ‘prohibit, or put some limitation on, the making or content of political communications’.⁹⁹

Crennan, Kiefel and Bell concluded in *Monis* that s 471.12 did place an effective burden on political discourse.¹⁰⁰ Their Honours accordingly focused on legal arguments relating to whether s 471.12 met the test in the second limb of *Lange*. The three judges considered s 471.12 valid because it did not ‘go too far’ in protecting people from receiving unsolicited offensive material in their ‘personal domain’.¹⁰¹

⁹⁵ In *Coleman v Power* (2004) 220 CLR 1 the judges on the High Court reached different conclusions about whether offensive communications can be political. McHugh J observed that ‘insults are a legitimate part of the political discussion protected by the *Constitution* [and] an unqualified prohibition on their use cannot be justified as compatible with the constitutional freedom’: 54 [105]. In the same case, Gummow and Hayne JJ, wrote that ‘[i]nsult and invective have been employed in political communication at least since the time of Demosthenes’: at 78 [197]. Kirby J noted that Australian political discourse frequently involved ‘insult and emotion, calumny and invective’: at 91 [239]. Gleeson CJ stated that the language was ‘not party political, and ... had nothing to do with any laws, or government policy’: at 31 [30]. Heydon J, who was more prosaic, said it was unnecessary for people to use *insulting words* to communicate about government and political matters: 30 [28].

⁹⁶ *Monis* (2013) 295 ALR 259, 281. French CJ, 282, also noted the broad application of the prohibition. His Honour also held that s 471.12 did *not* serve a legitimate end in a manner compatible with the constitutionally prescribed system of government and did not satisfy the second limb of *Lange*. Heydon J generally agreed with the reasoning of French CJ (and with the orders of French CJ and Hayne J).

⁹⁷ *Ibid* 315. In relation to the second limb of *Lange*, Hayne J did not believe that the end or object pursued by s 471.12 *was* a legitimate end. His Honour held that the provision did no more than regulate the giving of offence.

⁹⁸ *Ibid* 291.

⁹⁹ *Ibid* 290. See French CJ, 280, and Crennan, Kiefel and Bell JJ, 344. Crennan, Kiefel and Bell JJ, 344, noted, however, that an ‘effect upon political communication which is so slight as to be inconsequential may not require an affirmative answer to the first limb enquiry’.

¹⁰⁰ *Ibid* 344.

¹⁰¹ *Ibid* 340 (Crennan, Kiefel and Bell JJ).

Noting that s 471.12 was not *directed* to limiting political communications and that it only incidentally burdened such speech (citing, inter alia, *Wotton* for the relevance of this point¹⁰²), Crennan, Kiefel and Bell JJ reasoned that the statute's objective of protecting citizens from receiving offensive correspondence in the home or workplace was compatible with the implied freedom.¹⁰³ Their Honours also noted that s 471.12 limited communications of a *seriously* offensive kind.¹⁰⁴ Crennan, Kiefel and Bell JJ thus found that s 471.12 did not impermissibly burden freedom of political speech and held that the appeals should be dismissed.¹⁰⁵

III HOW FAR DOES THE IMPLIED FREEDOM GO?

A Lange's Strengths and Limitations

Lange remains the High Court's authoritative statement of the law relating to the implied freedom. It can be reasoned from *Lange* that the scope of the implied freedom is not limited to the speech of individuals or non corporations. The implied freedom may, for example, protect political speech by media organisations who are reporting or commenting on others' views, or who may be expressing their own political opinions (or those of their journalists, either freelance or employed).¹⁰⁶ By the same reasoning, the implied freedom *may* apply to speech of the delegates or leaders of organisations (perhaps even religious leaders) who — directly or indirectly; explicitly or implicitly — communicate about matters relevant to representative and responsible government on behalf of their people.

The protective scope of the implied freedom is determined by reference to what the burden does on political expression; its scope does not depend on how a person might want to characterise a particular communication.¹⁰⁷ This approach is consistent with the implied freedom not conferring constitutional *rights* to freedom

¹⁰² *Wotton* (2012) 285 ALR 1; *ibid* 344.

¹⁰³ *Ibid* 346.

¹⁰⁴ *Ibid* 346–7.

¹⁰⁵ *Ibid* 347.

¹⁰⁶ *Lange* concerned speech by the Australian Broadcasting Corporation.

¹⁰⁷ See *APLA* (2005) 224 CLR 322, 451 (Hayne J) cited with approval in *Monis* (2013) 295 ALR 259, 280 (French CJ); *Sunol v Collier [No 2]* [2012] NSWCA 44 (22 March 2012) [24] (Bathurst CJ) who nevertheless noted that 'the acts complained of may be of assistance in identifying the type of publications or speech which would generally fall within the challenged sections'. I note, briefly, that there is awkwardness in the suggestion that the inquiry under the first limb of *Lange* should be assessed by reference to the general effect of a burden on political speech. It is rare for courts to hear evidence on how a burden generally affects political speech (and the High Court would almost never hear direct evidence on this question given its appellate role in most implied freedom matters). It is also not clear how this approach would work in practice if there is a burden on the specific speech of an individual (such as might, say, occur in relation to a contempt of Parliament allegation).

of expression.¹⁰⁸ It can nevertheless be expected that the courts will consider the scope of the implied freedom by reference to *how* the relevant burden affects expression (including, perhaps most obviously, with respect to the person who is alleging the limitation on their political speech).¹⁰⁹ It is also now clear, if there were any doubt, that protected political speech may comprise offensive discourse¹¹⁰ and/or views that some consider lie at the fringe of political discussion.¹¹¹ A related point is that a law may burden political speech in its terms, operation or effect even if the law burdens only *some* political discourse.¹¹² Burdened political discourse *may* arguably include expression of religious viewpoints — *perhaps* even evangelistic messages — *if* the speech relates to federal politics.¹¹³

Lange, however, offers courts no bright line test in order to determine whether speech is political. Indeed, in *Lange*, the High Court sanguinely hinted that even *it* might be uncertain about the scope of the implied freedom, noting that ‘[w]hatever the scope of the implications arising from responsible government ... , those implications cannot be confined to election periods relating to the federal Parliament’.¹¹⁴ This suggests legal uncertainty about what kind of speech the implied freedom protects. This is to say nothing of the difficulties litigants, experts or commentators might face in predicting how courts might assess the *second* limb of *Lange* (as varied by McHugh J in *Coleman v Power*¹¹⁵) — which requires assessing whether the relevant legislation is reasonably appropriate and adapted to serve a legitimate

¹⁰⁸ *Lange* (1997) 189 CLR 520, 559.

¹⁰⁹ See *Sunol v Collier [No 2]* [2012] NSWCA 44 (22 March 2012) [24] (Bathurst CJ).

¹¹⁰ As noted, all six High Court judges in *Monis* concluded that s 471.12 of the *Criminal Code* burdened the defendant’s alleged offensive political speech.

¹¹¹ See *Monis* (2013) 295 ALR 259, 281 (French CJ), 293 (Hayne J). Cf *Coleman v Power* (2004) 220 CLR 1, 30 (Gleeson CJ).

¹¹² *Monis* (2013) 295 ALR 259, 290 (Hayne J), 280 (French CJ), 344 (Crennan, Kiefel and Bell JJ). Crennan, Kiefel and Bell JJ, 344, noted, however, that an ‘effect upon political communication which is so slight as to be inconsequential may not require an affirmative answer to the first limb enquiry’.

¹¹³ *A-G v Adelaide* (2013) 295 ALR 197 is *arguably* authority for this proposition. None of the majority judgments in *A-G v Adelaide* expressed doubt about whether ‘preaching’ could be constitutionally protected speech (Heydon J, the only dissenting judge in *A-G v Adelaide*, did not dissent on *this* point). That said, as noted, the South Australian Solicitor General *conceded* that the by-law burdened political speech. On the question of whether Caleb Corneloup (President of Street Church) or, in the case of *Monis*, Man Haron Monis (Sheikh Monis), were religious leaders, it is not clear to me whether they were, or are, religious leaders in the sense that they have ‘sizeable’ groups of followers — a working criterion for religious leader as I suggested at above n 1. In the *Monis* decision, the High Court did not mention whether Man Haron Monis’ speech may have been religiously motivated and it is not clear to me whether, though a Sheikh, Monis had or has a sizeable group of followers.

¹¹⁴ *Lange* (1997) 189 CLR 520, 561 (emphasis added).

¹¹⁵ (2004) 220 CLR 1.

end in a manner compatible with the maintenance of representative and responsible government provided for in the *Constitution*.¹¹⁶

Another challenge in using *Lange* (and other High Court cases) to determine the protective scope of the implied freedom is that *Lange* purports to ground political speech in the text and structure of the *Commonwealth Constitution*.¹¹⁷ While sourcing the implied freedom in the text of the *Constitution* does help to clarify its potential protective scope, the principle raises two important points. First, in several of its later decisions the High Court has sought to explain how speech affected by state laws relates to the *federal* system of government,¹¹⁸ but the High Court has not always done so. For example, in *Levy* (decided in 1997), there was no obvious political connection between Victorian regulations prohibiting duck hunting and federal political affairs (and nor did the High Court seek to explain why there might be one). Secondly, while the implied freedom is said to be based in the words of the *Commonwealth Constitution*, there is sparse guidance for defining political speech in the text of the *Constitution*. It is unlikely that the words in the *Constitution* alone will ever provide a *complete* guide to assist courts in identifying the scope of the implied freedom.

There also remains a tension between the scope of the implied freedom as described in the joint judgment in *Theophanous* and the High Court's later decision in *Lange*. In *Theophanous*, as noted, Mason CJ, Toohey and Gaudron JJ indicated that political discussion includes all speech that is relevant to the development of public opinion on the whole range of issues which an intelligent citizen should think about.¹¹⁹ At the time of the High Court's decision in *Theophanous*, the High Court had not universally agreed upon the proper basis for the implied freedom. While the High Court in *Lange* did not expressly overrule Mason CJ, Toohey and Gaudron J's judgment

¹¹⁶ The High Court affirmed this approach in *Aid/Watch* (2010) 272 ALR 417. The doctrines of representative and responsible government are, as suggested, the constitutional touchstones for defining political speech under the first limb of *Lange*. These doctrines are, however, also key elements of the criteria under the second limb of *Lange* for determining whether a law is reasonably appropriate and adapted to serve a legitimate end in a manner compatible with the maintenance of representative and responsible government provided for in the *Constitution*.

¹¹⁷ That is, unless the implied freedom arises under a relevant state constitution. See *Muldowney v South Australia* (1996) 136 ALR 18, 23. Brennan CJ noted that the Solicitor-General for South Australia conceded that there was a constitutionally entrenched limitation upon state legislative power 'in like manner to the *Commonwealth Constitution*' and that that limitation precluded interference by an ordinary law with freedom of discussion about political affairs (quotation marks in Brennan CJ's judgment referring to words of Solicitor-General).

¹¹⁸ See, eg, French CJ in *Hogan v Hinch* (2011) 275 ALR 408, 426, and the High Court's joint judgment in *Wotton* (2012) 285 ALR 1, 9.

¹¹⁹ *Theophanous* (1994) 182 CLR 104, 123.

in *Theophanous*, it is arguable that the subsequent clarification of the law in *Lange* narrowed its scope.¹²⁰

Yet, despite their differences, both *Theophanous* and *Lange* indicate that the implied freedom includes speech about representative government *and* responsible government. For example, the *Theophanous* joint judgment describes political speech in terms of the government's 'conduct' and 'policies' (executive actions) as well as the 'fitness for office of government ... and those seeking public office' (voting choices).¹²¹ Similarly, the High Court in *Lange* held that the protective scope of the implied freedom includes communications that are relevant to people making informed voting choices¹²² and the actions (presumably the inactions too) of the Commonwealth Government and its ministers, including through the ministers' departments, statutory authorities and public utilities.¹²³ The Court also observed in *Lange* that 'the conduct of the executive branch is not confined to ministers and the public service'.¹²⁴ While it would be stretching the interpretation of these words to contend that *Lange* recognised speech about, say, lobbyists (including religious leaders or leaders of religious lobby groups), as part of political discussion,¹²⁵ it would seem from this phrase in *Lange* that the High Court *did not* have a narrow conception of executive responsibility (ie, speech about executive responsibility under the implied freedom at least reaches *beyond* only discussion of ministers and public servants).

In subtly different ways, each joint judgment — *Theophanous* and *Lange* — describes the implied freedom in terms of speech about voting choices and executive responsibility. While there are obvious differences in the decisions (reflecting the High Court's evolving reasoning in this field of law), this observation hints at there being possibly fewer differences between the *Lange* and *Theophanous* judgments than might initially meet the eye. It is even possible that a generous view of the kind of political speech that is relevant to voting choices and responsible government might reveal additional concentricity between the *Theophanous* and *Lange* judgments. For example, it might be said that each judgment supports the view that the protective scope of the implied freedom extends to matters over which the government *does not yet have* responsibility. Additionally, if political discourse is viewed as dynamic, and if the Court accepts that the development of public opinion can, over time, influence voting choices or even

¹²⁰ See *Brown v Classification Review Board* (1998) 154 ALR 67, 85–6. Heerey J suggested that *Lange* necessarily qualified the definition of political speech that Mason CJ, Toohey and Gaudron JJ proposed in *Theophanous*. See also *Sunol v Collier [No 2]* [2012] NSWCA 44 (22 March 2012) [84] where Basten JA expressed doubt about whether *Theophanous* 'accurately reflects an immunity limited to that which is necessarily implied from the text and structure of the *Constitution*' as required by *Lange*.

¹²¹ *Theophanous* (1994) 182 CLR 104, 123.

¹²² *Lange* (1997) 189 CLR 520, 559.

¹²³ *Ibid.*

¹²⁴ *Ibid.*

¹²⁵ The joint judgment in *Theophanous*, by contrast, would likely support such an interpretation.

the electorate's expectations of standards of responsible government, then it may be possible to, in some way, *further* reconcile the two judgments.

B *What does Speech about Representative and Responsible Government Mean?*

What then, does it *mean* for the implied freedom to protect speech that is relevant to representative and responsible government? In the following passages, I try to answer this question in seven points. I begin by describing the interrelationships between representative and responsible government. I then consider the significance of voting preferences with respect to speech that might fall within the protective scope of the implied freedom. I later explore the importance of temporal considerations in relation to the concepts of representative and responsible government and, in doing so, briefly return to *Lange*. The objective in doing so is to try to identify more clearly whether political speech as the High Court described it in *Lange* can include speech by or about religious leaders.

First, the touchstone of *responsible* government is that the executive must answer to the Commonwealth Parliament. By contrast, *representative* government refers to a system whereby elected representatives are accountable to the people who vote for them. In theory, representative government and responsible government are *distinct* political frameworks and yet there may a degree of interconnectedness between the two systems. A government's actions when it is in office — its actions and inactions — can (and do) affect voter choices. Additionally, speech about the kinds of things a government *does* when governing (and also what it does not do or *perhaps what it should* do) can influence voters' perceptions of the government's effectiveness. This, in turn, may affect voters' decisions at the ballot box. It may, as a result, be difficult to distinguish speech that is relevant to responsible government from speech that pertains to representative government and vice versa.

Secondly, arguably the most salient guidance in relation to any system of government in the *Constitution* is in ss 7 and 24. Notwithstanding the possible above-mentioned overlaps between representative government and responsible government, ss 7 and 24 tend to focus on voters' choices and reveal that *representative* government is, at least from the text of the *Constitution*, a governmental system that underpins the constitutional arrangements. Using the expression 'directly chosen by the people', ss 7 and 24 indicate that voters have a direct *choice* — ie, that the voting public can 'choose' between different Commonwealth parliamentary candidates — conversely, voters can elect to *not* choose political candidates who lack sufficient appeal. Dawson J in *ACTV* noted the significance of ss 7 and 24 of the *Constitution* and the importance of voter choices for the implied freedom.¹²⁶ Hayne J also referred to those sections (and ss 64 and 128) in *Monis*.¹²⁷ Expression that is relevant to

¹²⁶ (1992) 177 CLR 106, 186–7.

¹²⁷ (2013) 295 ALR 259, 288. Justice Hayne observed, further, that 'those who are elected as members of the parliament and those who are appointed as ministers of state are necessarily accountable to "the people"'. (His Honour thus implicitly distinguished representative government from responsible government while highlighting the accountability of political representatives to the electorate).

Commonwealth voting choices will be, and will likely remain (in the absence of constitutional change), a cornerstone of political expression.

Thirdly, different voters will likely have different wants in relation to acquiring and passing on information that is relevant to their voting choices. People do not have homogenous desires for political information. The kind of material that an individual regards as relevant to their voting decisions is that which allows the person to exercise a genuine choice. Voting is a matter of *personal* preference and the word ‘direct’ in ss 7 and 24 in relation to that voting choice supports this view.¹²⁸ Disinterested or disaffected voters may need little evidence to inform their voting decisions. Others, by contrast, may seek as much material as possible. Some voters may be primarily concerned with the political matters of the day and their political concerns may be short term. Others, by comparison, may take a longer term view of political affairs. For some individuals, the kind of expression that is relevant to their own ballot box choices and to their perceptions of the government’s actions could be about a topic that is barely spoken of, except for, say, in academic journals or in their local church. For others, the topic of primary political interest may be at the forefront of contemporary parliamentary debate. At the ballot box, one voter could be interested in the personal background, professional affiliations or lived ethical standards of a political candidate in the voter’s own electorate or elsewhere. Another voter may, by contrast, be concerned about the policies or practices of an entire political party. A voter’s choice about whether to elect a particular electoral candidate could be influenced by considerations outside his/her electorate or in other tiers of government. A voter’s decision may be influenced by a policy in relation to which a government does *not* want to take responsibility; or over which governments tend to share responsibilities.

Some voters may be influenced by the views or teachings of religious leaders and may search for political parties or candidates whose views resonate with those teachings. Depending on voters’ dispositions for acquiring political information and their perceptions of the importance of political matters (and the significance of politicians to them), for some voters, perhaps even *many*, politics will be about government over the long term — leadership, welfare, health, economic wellbeing and the good of society. For others, ballot box decisions may involve little more than the whimsical filling out of ballot papers at the local school on a Saturday every few years and may have no more political significance than that. Whoever the voter may be, it is inherent in them making an *informed* voting choice that they are able to take into account such *information as they see as relevant to their choice about who to elect to political office*.

Fourthly, a wide range of matters fall within Commonwealth government executive responsibility. Speech about executive responsibility could include commentary about the decisions (or words or speeches) of ministers or the actions of the Prime Minister and also — as the High Court wrote in *Lange* — discourse about the

¹²⁸ ‘Direct’ implies that the voter’s choice at the ballot box is paramount, in contrast to the views of, say, lobby groups or even political parties.

actions (or inactions) of ministers' departments and statutory authorities. Speech that is relevant to responsible government could comprise expression about the government's actions in creating legislation, executing policies or implementing programs. Political speech could be about what the government refuses to do, or does not do. Speech about responsible government may include expression about possible perceived *religious* objectives the government may prosecute. Such religious objectives might, for example, comprise opposing legalisation of euthanasia or — a more contentious suggestion — opposing onshore processing of asylum seekers because of implicit objections to the possible settlement of Muslims in Australia.

Fifthly, over time, executive responsibility can switch between tiers of government and government responsibilities can change. Governments often jointly administer policies (eg, there are joint Commonwealth and state roles in relation to law enforcement, aspects of taxation, education, and health). Hearings and findings of commissions of inquiry may highlight connections between members of political parties at different tiers of government. Some politicians may switch from state government positions to the federal government (or from local government to state government roles). There may be a morphing of executive responsibilities and a blending of perceived accountabilities for legislative and executive decisions between Commonwealth, state, or even local branches of government. The changing nature of these governmental responsibilities may make it difficult — especially over time — to identify any single tier of government as having responsibility for a particular field.

Sixthly, political speech is not a static concept nor is it fixed in time. A generous view of political speech may facilitate discourse about nascent political matters. French CJ seemed to accept as much in *Hogan v Hinch* when his Honour described the implied freedom as applying to expression about 'matters *potentially within the purview of government* (emphasis added).¹²⁹ Political speech may alter depending on the composition of Parliament. Consider climate change. Perhaps only two decades ago, notwithstanding that the Commonwealth government already had a well-established role in environmental law, many Australian voters may *not* have conceived of climate change as an area of significant government responsibility, let alone a topic that might be relevant to people's voting choices. Yet, scientific research and academic debate in this field — *speech* and *expression* — arguably drove political leaders in the Western world (including Australia) to recognise climate change as a bellwether topic for progressive Parliaments. To *not* recognise climate change as within the scope of the implied freedom in, say, 1990 could have denied the Australian electorate the ability to communicate about a matter that would emerge little more than two decades later as an arguably critical consideration in influencing voting choices for Australians. The same might be said of asylum seeker policy in Australia; Islam; poker machines; and real-time advertising of online bookmakers' odds on the outcomes of broadcast sporting

¹²⁹ (2011) 275 ALR 408, 426.

games — that is, these matters *became* political topics without necessarily being inherently so.¹³⁰

Seventhly, voting choices may traverse parliamentary terms. Clearly, the actions of a government *while it is governing*, can stimulate public debate and affect public sentiment (and voting choices) for long periods, again hinting (though the concepts are distinct) at the lockstep nature of voting choices and executive responsibility. Government budgetary decisions in one term can affect perceptions about the economic competence of later governments. Conversely, ostensibly ‘old’ political topics — for example, the government’s role in previous World Wars or governments’ treatment of the Stolen Generations to take just two examples — can *remain* political, affecting peoples’ lives (and influencing their voting choices) for decades. Voters may choose political parties and independent MPs on their *long-term* records. Political topics resurface as do former political leaders. Even *retired* political figures can exert strong influence amongst voters beyond their parliamentary careers.

Returning to *Lange* — beyond saying the implied freedom did not only apply during election periods, the High Court did not restrict the scope of the implied freedom to any time frame. The High Court in *Lange* also acknowledged that most of the information that people need to make an informed voting choice exists *between the holding of one and the calling of the next election* — meaning, a government’s actions in government can influence *voting* choices at *future* elections.¹³¹ Similarly, Heydon J in *Coleman v Power* described the implied freedom as protecting expression ‘which might lead to voter appreciation of the available alternatives’, implying that voters make future ballot box choices based on how candidates perform after being elected to office.¹³²

The High Court said in *Lange* that the information that is relevant to responsible government arises ‘throughout the life of a federal Parliament’.¹³³ This does not suggest a conservative approach to the temporal aspects of speech about responsible government. Similarly, in *Hogan v Hinch*, French CJ said political matters are ‘not limited to matters concerning the current functioning of government’.¹³⁴ People may (and do) remember the actions of their governments — fondly or otherwise — after those governments have been elected out of office. *Historic* acts of government can permeate peoples’ perceptions of *future* governments — in positive and negative ways — for years and perhaps even decades. By contrast, younger voters may have naive, though genuine, expectations of their government and may have their own perceptions of what

¹³⁰ Poker machines became a political topic for the Commonwealth as a result of Independent MP Andrew Wilkie requiring, in 2009, that the Gillard-led Labor Government support voluntary poker machine pre-commitment reforms to help ‘problem’ poker machine users in return for Mr Wilkie giving then Prime Minister Gillard his parliamentary vote on crucial supply bills.

¹³¹ (1997) 189 CLR 520, 561. The High Court thus rejected McHugh J’s suggestion in *ACTV* (1992) 177 CLR 106, 277 that the implied freedom only applies during election periods.

¹³² (2004) 220 CLR 1, 126.

¹³³ (1997) 189 CLR 520, 561.

¹³⁴ (2011) 275 ALR 408, 426.

is relevantly political in the short term or longer run. Emphasising the non transitory nature of executive responsibilities, ministers in previous state and federal governments can be called to account for their actions before later parliaments. Committees regularly review the actions and activities of former governments, ministers and departments. Conversely, House of Representatives and Senate committees often undertake inquiries to better inform the policies of current and *future* governments.

Given these considerations, it would be unrealistic to propose *any* rigid time frame for speech about responsible or representative government. I suggest there is another reason for the courts to avoid a rigid, static or short-term approach to deciding whether speech falls within the protective scope of the implied freedom. As I have hinted, topics can become *political* if people have sufficient freedom (and time) to discuss them. A less than generous approach to interpreting the temporal aspects of the implied freedom — in effect, not granting topics the opportunity to grow into political subjects in the longer run — could deny Australian voters the opportunity to freely communicate about nascent political subjects that may, over time, *become* important topics for governments if there is enough confidence to discuss them.

IV ILLUSTRATING THE LEGAL PRINCIPLES: SPEECH BY AND ABOUT RELIGIOUS LEADERS

I now briefly analyse whether speech by or about religious leaders might fall within the first limb of *Lange*. Two examples are considered. As suggested, the purpose of this study is not to review any specific burden that might apply to the speech. By reviewing these examples, I consider whether, if burdened, speech by or about religious leaders could fall within the protective scope of the implied freedom (ie, the first limb of *Lange*).

Consider the former Australia Mufti, Sheikh Taj El-Din Hamid Hilaly. In 2006, the Sheikh was reported to have said in a sermon that women who do not wear a hijab are like ‘uncovered meat’ (to whom men might be attracted sexually, as would a cat).¹³⁵ In context, the Sheikh’s comments were about criminal sentencing for young Muslim sexual offenders. Specifically, the Sheikh’s sermon referred to Sydney gang rapes and included the following passage (as reportedly translated from Arabic to English):

¹³⁵ In an edited English translation prepared by SBS, the Sheikh is reported to have said: ‘If you take uncovered meat and put it on the street, on the pavement, in a garden, in a park or in the backyard, without a cover and the cats eat it, is it the fault of the cat or the uncovered meat? The uncovered meat is the problem. If the meat was covered, the cats wouldn’t roam around it. If the meat is inside the fridge, they won’t get it. If the meat was in the fridge and it (the cat) smelled it, it can bang its head as much as it wants, but it’s no use. If the woman is in her boudoir, in her house and if she’s wearing the veil and if she shows modesty, disasters don’t happen’; see Dalia Mattar (SBS Translator), ‘What Sheikh al-Hilaly Said’, *The Australian* (online), 27 October 2006 <<http://www.news.com.au/breaking-news/what-sheik-al-hilaly-said/story-e6frfkp9-111112425700>>.

when it comes to adultery, it's 90 percent the woman's responsibility. Why? Because a woman owns the weapon of seduction. It's she who takes off her clothes, shortens them, flirts, puts on make-up and powder and takes to the streets ... It's she who shortens, raises and lowers. Then it's a look, a smile, a conversation, a greeting, a talk, a date, a meeting, a crime, then Long Bay Jail.¹³⁶

The references to 'a crime' and 'Long Bay Jail' were generally understood to mean the arrest and sentencing of Bilal Skaf, a young Muslim convicted of sexual assault.¹³⁷ Skaf was the leader of a gang of up to 14 Lebanese Muslim men who raped four women in Sydney in August 2000.

Arrested in November 2000, sent to trial, later convicted and sentenced to 55 years imprisonment (his sentence was later reduced on appeal to 40 years jail), Skaf's imprisonment became a political topic. For example, in 2003, the NSW Opposition Justice spokesman, John Ryan, accused the then NSW Justice Minister, John Hatzistergos, of being complicit in releasing cartoons that Skaf drew in prison (one depicted Skaf's fiancé being pack raped). Ryan claimed Hatzistergos allowed the pictures to be released in order to 'enhance [the government's] law and order credentials' for the purpose of showing off to the public the government's 'Super Max' jail.¹³⁸ Hatzistergos denied the allegation.

The Sheikh's audience — men in a Lakemba mosque during *Ramadan* — would likely have understood the Sheikh's message to be about prison sentences for Muslim men convicted of sexual assault. It might be said that the Mufti's comments were made in a private setting of religious worship — a Lakemba mosque — and the speech lacked any quality of political expression because of the separation of church and state (or, more accurately, mosque and state) and that 'mosque speech' cannot be political speech. I am not, however, aware of any constitutional principle preventing political speech including speech in mosques.

The reported aspects of the Mufti's sermon constituted speech about the courts' sentencing of male Muslim rapists; it was not speech about the (Commonwealth) executive nor was it particularly relevant to (Commonwealth) voting choices.¹³⁹

¹³⁶ Ibid. Reported translation by Dalia Mattar.

¹³⁷ For example, when referring to the Mufti's sermon in 2006, the then Commonwealth Treasurer Peter Costello was in no doubt that Sheikh Hilali was referring to the case of Bilal Skaf. See Interview with Alan Jones (31 October 2006) <<http://www.treasurer.gov.au/tsr/content/transcripts/2006/159.asp>>.

¹³⁸ See Stephen Gibbs, 'Skaf Stories Part of a Freak Show: Libs', *Sydney Morning Herald* (online), 26 July 2003, <<http://www.smh.com.au/articles/2003/07/25/1059084204009.html>>.

¹³⁹ See McHugh J in *APLA* (2005) 224 CLR 322, 361, who was in no doubt that the implied freedom would protect communications about judicial officers and courts concerning matters arising from legislative acts or executive decisions or omissions. By contrast, on Spigelman CJ's reasoning in *John Fairfax Publications Pty Ltd v A-G (NSW)* (2000) 181 ALR 694, 709, the implied freedom would not protect such expression.

In my view, if it were burdened, the Mufti's speech was *insufficiently* relevant to representative and responsible government — particularly in relation to the *Commonwealth* Parliament — to fall within the protective scope of the implied freedom.¹⁴⁰

While I believe that the Mufti's speech was not relevantly political, speech *about the Mufti's sermon could* still fall within the protective boundaries of the implied freedom if *that* speech, assuming it was burdened, had a sufficient connection with voting choices or executive responsibility in respect of the Commonwealth Parliament. The Mufti's statements generated a vast amount of commentary amongst senior state and federal politicians, all critical of the Mufti.¹⁴¹ The fact that so many senior politicians (state and federal) contributed to the debate about the reported aspects of the Mufti's sermon perhaps points to his comments touching on topics of political interest (ie, *the status of women or Islamic extremism*).

A senior Sydney Anglican — Dean Philip Jensen — spoke about the Sheikh's sermon and the Dean's comments bring an interesting perspective to the implied freedom and the Mufti's statements. The Dean's comments also raise the question of whether one religious leader's comments about *another* religious leader's statements could fall within the protective boundaries of the implied freedom. Dean Jensen claimed that the Mufti's sermon was about *adultery* and the media commentary failed to adequately appreciate the Sheikh's culture. Dean Jensen said:

The Sheik's comments were about the difference between men and women. The sermon was about adultery. It was about a gender distinction made in Quranic judgement between theft and adultery ...

The biggest failed shibboleth of the week was the separation of church and state. Political leaders from both sides of the house in both state and federal parliaments have pressured the Muslim community to sack a man for what he preached in a religious service. Secularists have been notably quiet on this issue of separation of State and Church. *They have written and spoken out against the Prime Minister's plan to subsidize chaplains in schools but not in his outspoken pressure to remove the Sheik from his post* (emphasis added).¹⁴²

¹⁴⁰ See *APLA* (2005) 224 CLR 322, 351 (Gleeson CJ and Heydon J), 362 (McHugh J); 450 (Callinan J).

¹⁴¹ The political figures critical of the Mufti included the then Prime Minister, John Howard and the then Treasurer, Peter Costello. There were explicit demands for the Mufti's resignation from the then Shadow Foreign Affairs Minister Kevin Rudd; the then NSW Premier, Morris Iemma; the then Victorian Premier Steve Bracks; and the then Federal Sex Discrimination Commissioner, Pru Goward (later to become NSW State Member for the Seat of Goulburn and who recommended the Sheikh be deported for allegedly inciting men to rape). The then NSW Minister for Emergency Services, Nathan Rees (later a NSW Labor Premier) called the Sheikh a 'serial boofhead'.

¹⁴² Phillip Jensen, *Sheikh Hilali* (2 November 2006) <<http://phillipjensen.com/articles/sheik-hilali2/>>.

The emphasised passage carries implicit criticism of those people who opposed then Prime Minister John Howard's policy of supporting funding for school chaplains. This was a controversial topic for the then Coalition government (and for Labor governments afterwards) — perhaps hinting at a government policy merging church and state. What is more, the chaplaincy program became the subject of a High Court challenge against the Commonwealth government.¹⁴³ The implication of the Dean's paper is that advocates of church-state separation were hypocrites for criticising Prime Minister Howard's support for funding school chaplains while not also objecting to the Prime Minister's public urges to remove the Sheikh from his religious position. The Dean's expressed support for the Howard government's policy of funding school chaplains, was, in my view, relevant to Sydney Anglicans' voting choices and/or their understanding of the actions of the Commonwealth executive and this speech, assuming there was a relevant burden, would likely fall within the protective boundaries of the implied freedom (ie, under the first limb of *Lange*).

V CONCLUSION

Lange remains the authoritative High Court case on the scope of the implied freedom. According to *Lange*, the implied freedom applies to speech that facilitates informed voting choices *and* discourse which is relevant to the effectiveness of the government, including through its statutory authorities and public utilities, in governing for people who exercise those voting choices. By contrast, Mason CJ, Toohey and Gaudron J's joint judgment in *Theophanous* indicates that the implied freedom applies to speech about the political views and public conduct of all persons involved in activities that have become the subject of political debate.

The *Theophanous* view of the scope of the implied freedom appears broader than what the High Court decided in *Lange*. The earlier joint judgment would also appear to more readily recognise religious leaders' contributions to public debate as political speech. I have contended, however, that a generous interpretation of the High Court's description of the implied freedom in *Lange* could result in the differences between *Lange* and *Theophanous* being less stark than otherwise may appear. It is also possible that, over time, speech which is relevant to the development of public opinion can *become* the kind of speech that is important for peoples' voting choices or their perceptions of executive responsibility and this may again imply more similarities between *Theophanous* and *Lange* than might otherwise initially seem apparent.

I have also suggested that a generous temporal view of political speech could protect discussion about topics that are of emerging political relevance and this may allow discourse about those subjects to grow or even flourish. Political speech *may* thus include the contributions of religious leaders to public debate even if the subject matter is of nascent relevance to representative and responsible government (or, by contrast, is apparently of primarily historic relevance).

¹⁴³ See *Williams v Commonwealth* (2012) 288 ALR 410.

It is, accordingly, possible to answer the questions I posed at the start of this article. First, political speech is speech or other expression that is relevant to voting choices or executive responsibility in relation to the Commonwealth Parliament. Political speech can include speech about state government matters if these are relevant to voting choices or executive responsibility in relation to the Commonwealth Parliament. Secondly, the implied freedom may protect speech by and about religious leaders from a burden if the speech is political speech as defined above (and the plaintiff can also show that the relevant burden does not fall within the second limb of *Lange*, as varied by McHugh J in *Coleman v Power*).

These answers, I suggest, reflect the minimum limits of political speech. Based on the *Theophanous* view of political speech, though perhaps not inconsistent with *Lange*, the implied freedom could protect speech by or about religious leaders who seek to influence government decision making (eg, by lobbying). It might also be argued that *A-G v Adelaide*¹⁴⁴ supports the view that political discourse can include expression of religious viewpoints — perhaps even evangelism — if the speech relates in some way to federal politics.¹⁴⁵ The above responses may have important implications for the speech of religious leaders in Australia who contribute to public debate (and for those who participate in discourse about them), for, on this reasoning, that kind of speech *may* fall within the protective scope of the implied freedom (ie, the first limb of *Lange*).

¹⁴⁴ (2013) 295 ALR 197.

¹⁴⁵ As suggested, because of the South Australian Solicitor General's concession in *A-G v Adelaide* that the by-law restricted political expression, the authority of the case may be somewhat limited in respect of the suggestion that political speech may include expression of religious viewpoints, including evangelism. Significantly, the High Court found in that case that the by-law *was* valid (as a matter of statutory construction and under the second limb of *Lange*).

*Emily Bell**

NEW SOUTH WALES v KABLE (2013) 298 ALR 144

I INTRODUCTION

In *New South Wales v Kable* (*'Kable [No 2]'*),¹ the High Court of Australia considered whether an order made under the *Community Protection Act 1994* (NSW) permitted the lawful detention of Gregory Wayne Kable when the judicial nature of that order had been questioned. This case note compares the High Court's decision with that in *Kable v Director of Public Prosecutions (NSW)* (*'Kable [No 1]'*)² in light of the difficulties faced by appellate courts when confronted with apparently contradictory and restrictive authorities.

II BACKGROUND

A The Original Proceedings: Kable [No 1]

On 23 February 1995, Levine J of the Supreme Court of New South Wales ordered that Mr Kable be detained in custody for six months.³ No criminal trial took place before his Honour; the order was instead made under s 9 of the *Community Protection Act 1994* (NSW) (*'the CPA'*). This unique piece of legislation permitted *'the preventive detention ... of Gregory Wayne Kable'*⁴ upon application by the Director of Public Prosecutions,⁵ provided the Supreme Court of New South Wales (*'the Supreme Court'*) was satisfied that Mr Kable was likely to commit a *'serious act of violence'*⁶ and that he posed a prospective threat to the community.⁷

Mr Kable appealed unsuccessfully⁸ to the New South Wales Court of Appeal.⁹ He was ultimately granted special leave to appeal to the High Court. In separate judgments delivered long after the six month period of imprisonment had expired, the majority of Toohey, Gaudron, McHugh and Gummow JJ upheld Mr Kable's

* Law Student, Adelaide Law School, The University of Adelaide.

¹ (2013) 298 ALR 144.

² (1996) 189 CLR 51.

³ *Ibid* 146.

⁴ *CPA* s 3(1).

⁵ *Ibid* s 5(1).

⁶ *Ibid* s 5(1)(a).

⁷ *Ibid* s 5(1).

⁸ *Ibid* s 25.

⁹ *Kable v DPP (NSW)* (1995) 36 NSWLR 374.

appeal.¹⁰ Their Honours found that the *CPA* was invalid because it bestowed a power upon the Supreme Court that was incompatible with its capability to exercise federal jurisdiction under Ch III of the *Commonwealth Constitution*.¹¹ The judgments of Gaudron, McHugh and Gummow JJ in particular focused on the non-judicial nature of the power exercised under the *CPA*, finding it to be more in line with an exercise of executive power.¹²

B *The Present Case at First Instance*

Following the High Court's decision, Mr Kable issued proceedings against the State of New South Wales ('the State') seeking damages for false imprisonment, and later adding claims for abuse of process and malicious prosecution.¹³ At first instance, Hoeben J found for the State.¹⁴

C *In the New South Wales Court of Appeal*

Mr Kable lodged a partially successful appeal to the New South Wales Court of Appeal ('the Court of Appeal'). The Court of Appeal found for Mr Kable on the issue of false imprisonment, holding that the order of Levine J did not provide lawful authority for Mr Kable's detention.¹⁵ Based on their interpretation of *Kable [No 1]*, the Court of Appeal perceived the detention order as 'an invalid non-judicial order'.¹⁶ In a particularly striking judgment, Allsop P concluded that in exercising the non-judicial power conferred by the *CPA*, the Supreme Court had stepped outside its role as a superior court of record. His Honour found that it essentially acted as an extension of the executive arm of government.¹⁷

D *In the High Court: Kable [No 2]*

The State was granted special leave to appeal to the High Court ('the Court'). The Court unanimously allowed the State's appeal and set aside the orders of the Court of Appeal, ordering that the appeal to that court be dismissed.¹⁸ The Court accepted that the *CPA* was an invalid law. However, the Court found that the preventive detention order made by Levine J provided an independent source of authority for Mr Kable's detention until it was set aside: on their Honours' interpretation, it was a judicial order made by a superior court.¹⁹

¹⁰ *Kable [No 1]* (1996) 189 CLR 51.

¹¹ *Ibid* 99 (Toohey J), 107 (Gaudron J), 124 (McHugh J), 144 (Gummow J).

¹² *Ibid* 106 (Gaudron J), 122 (McHugh J), 132 (Gummow J).

¹³ *Kable [No 2]* (2013) 298 ALR 144, 146.

¹⁴ *Kable v New South Wales* (2010) 203 A Crim R 66.

¹⁵ *Kable v New South Wales* (2012) 293 ALR 719, 759, 762.

¹⁶ *Ibid* 758 (Basten JA).

¹⁷ *Ibid* 722 (Allsop P).

¹⁸ *Kable [No 2]* (2013) 298 ALR 144, 155.

¹⁹ *Ibid* 151 (French CJ, Hayne, Crennan, Kiefel, Bell and Keane JJ), 163 (Gageler J).

The plurality of French CJ, Hayne, Crennan, Kiefel, Bell, and Keane JJ contradicted the Court of Appeal's interpretation of *Kable [No 1]*. Their Honours stated that the decision in *Kable [No 1]* was predicated upon the Supreme Court being required to act as a court while exercising a quasi-executive power that was incompatible with that role.²⁰ On this analysis, the problem with the *CPA* emanated from the *requirement* that the Supreme Court maintain its role as a superior court of record when making a preventive detention order.

Their Honours also placed significant weight on the process adhered to in the original proceedings. It was never disputed that the orders of Levine J and those of the Court of Appeal engaged and fell within the High Court's appellate jurisdiction.²¹ Their Honours explained that by allowing the appeal to the Court of Appeal and setting aside the order of Levine J in *Kable [No 1]*, the High Court treated Levine J's decision as if it were a judicial order.²²

The plurality distinguished the making of the detention order from cases where judges exercised executive power²³ by examining the manner in which the power was exercised by Levine J. They concluded that because the detention application was heard *inter partes*, some of the rules of evidence applied, and the decision was appealable, the order of Levine J was properly characterised as judicial.²⁴ Their Honours denied that Levine J's exercise of federal jurisdiction (in deciding that he *had* jurisdiction to make an order) and the exercise of power in making the detention order under the *CPA* required separate consideration.²⁵ The legal effect of Levine J's order derived not from the invalid power purportedly conferred by the *CPA*, but from its status as a judicial order of a superior court, valid until set aside even if made beyond jurisdiction.²⁶

In a separate judgment, Gageler J focused almost exclusively on the effect of the appeals in the original proceedings. The High Court heard the appeal in *Kable [No 1]* pursuant to s 73 of the *Commonwealth Constitution*. This only permitted the High Court to determine an appeal from 'a decision made in the exercise of judicial power'.²⁷ The High Court proceeded to set aside the orders of Levine J and ordered instead that the application under the *CPA* should be dismissed.²⁸ Because the orders of Levine J could only have been treated in that way in the exercise of judicial power, his Honour concluded that the Court of Appeal in the original proceedings must also have exercised judicial power in considering the orders of Levine J.

²⁰ Ibid 148.

²¹ Ibid 148–9; *Commonwealth Constitution* s 73(ii).

²² *Kable [No 2]* (2013) 298 ALR 144, 149, 151.

²³ See, eg, *Love v A-G (NSW)* (1990) 169 CLR 307, 321–2.

²⁴ *Kable [No 2]* (2013) 298 ALR 144, 151.

²⁵ Ibid 153–4.

²⁶ Ibid 153.

²⁷ Ibid 162.

²⁸ Ibid.

His Honour acknowledged that the Court of Appeal sometimes reviewed administrative decisions, but found no reason to treat the order of Levine J differently to the judicial order of the Court of Appeal: both orders were purportedly made under the *CPA*, which intended for them to be made in adversarial conditions.²⁹ Upon finding that the order of Levine J was a judicial order made by a superior court, Gageler J accepted that the order was valid until set aside.³⁰

III *KABLE [No 2]*: CLARIFICATION OR CONFUSION?

A Inconsistencies

The reasons in *Kable [No 2]* are somewhat difficult to reconcile with the majority's reasoning in *Kable [No 1]*. Certainly, the Court of Appeal could be excused for interpreting comments found in *Kable [No 1]* as declaring the order of Levine J an 'invalid non-judicial order'.³¹ Suggestions that the power was 'non-judicial'³² but was 'purely executive in nature'³³ would lead reasonable minds to that conclusion.

Further inconsistencies with *Kable [No 1]* are apparent in the Court's view of the very character of judicial power and process. In *Kable [No 2]*, the Court concluded that the exercise of power by Levine J bore many of the hallmarks of judicial process: the application was heard inter partes, some of the rules of evidence applied, and the decision involved adjudication upon the rights of the parties.³⁴ On the other hand, the majority in *Kable [No 1]* found that the way the power was to be exercised under the *CPA* was 'repugnant to the judicial process in a fundamental degree'³⁵ and represented 'the antithesis of the judicial process'.³⁶ Gaudron J could not accept that the decision to be made by Levine J involved the 'resolution of a [civil] dispute between contesting parties as to their respective legal rights and obligations'.³⁷ The decision bore no more resemblance to an assessment of criminal liability. Her Honour concluded that despite being dressed up as a legal proceeding, the process under the *CPA* involved guesswork based in part on ordinarily inadmissible evidence.³⁸

²⁹ Ibid.

³⁰ Ibid 158, 163.

³¹ *Kable v New South Wales* (2012) 293 ALR 719, 758.

³² *Kable [No 1]* (1996) 189 CLR 51, 132.

³³ Ibid 122.

³⁴ *Kable [No 2]* (2013) 298 ALR 144, 151 (French CJ, Hayne, Crennan, Kiefel, Bell and Keane JJ), 163 (Gageler J).

³⁵ *Kable [No 1]* (1996) 189 CLR 51, 132.

³⁶ Ibid 106.

³⁷ Ibid.

³⁸ Ibid 106–7.

B *The Effect on Appellate Courts*

In *Kable [No 2]*, the Court attempted to avoid an undesirable application of the principles outlined in *Kable [No 1]* without purporting to overrule the earlier decision. This approach creates problems for other appellate courts as they act in their role to not only interpret the law, but to gradually develop and clarify it.³⁹ Contradictory authorities slow development and make clarification virtually impossible.

For example, the Court of Appeal in these proceedings made a clear attempt to directly apply the principles outlined in *Kable [No 1]*. Although it is a role of the High Court to review the decisions of appellate courts, the decision in *Kable [No 2]* effectively took away the Court of Appeal's power to adjudicate the matter based on authority. In particular, the narrow reasoning employed by Gageler J took the matter out of the Court of Appeal's hands: upon noting that the appellate court in the original proceedings had accepted its task of reviewing the decision of Levine J, the Court of Appeal should not have embarked upon an application of the principles outlined in *Kable [No 1]*. Essentially, it was not permitted to conclude that the order of Levine J was anything other than a judicial order. The proliferation of this kind of reasoning creates a minefield for appellate courts endeavouring to interpret, apply, and build upon the highest common law authorities.

Going forward, the inconsistencies between *Kable [No 1]* and *Kable [No 2]* may continue to cause confusion. As highlighted above, the two decisions took opposing approaches to defining the character of judicial process. They assigned differing levels of importance to the indicators that judicial process had been followed, and interpreted the nature of the judicial process in different ways. However, both approaches appear to remain good law. Clarification of this issue is likely to be slow, barring a suggestion that parts of *Kable [No 1]* have been overruled.

C *An Alternative Solution*

There is no doubt that finding the State liable in tort for complying with a purportedly judicial order would have been an unsatisfactory result. Indeed, the plurality drew attention to what appeared to be policy reasons for denying Mr Kable relief in the circumstances: if the order of Levine J had no lasting legal effect until it was subjected to final review, the parties would be caught between their obligation to obey the order and the possibility of incurring tortious liability for doing so until final review occurred.⁴⁰ However, the Court could have solved this conundrum by acknowledging that lawful justification for false imprisonment included the enforcement of a superior court order that was *prima facie* valid. This argument was put forward by counsel for the State, but was not explored by the Court.⁴¹ The required

³⁹ J D Heydon, 'Judicial Activism and the Death of the Rule of Law' (2003) 23 *Australian Bar Review* 110, 124.

⁴⁰ *Kable [No 2]* (2013) 298 ALR 144, 154.

⁴¹ *Ibid.*

expansion of the scope of lawful justification would be small,⁴² and it would only be necessary to apply the expanded doctrine in limited circumstances where a detention order made by a superior court is found to lack validity.

IV CONCLUSION

Despite avoiding an undesirable outcome, the Court in *Kable [No 2]* failed to directly address the apparent inconsistencies between their decision and that in *Kable [No 1]*. Some aspects of the Court's reasoning were overly restrictive, while other aspects contradicted *Kable [No 1]* without purporting to disturb its authority. This makes the role of appellate courts difficult as they attempt to clarify and develop the common law. The Court's narrow reasoning has a direct effect on courts of appeal by curbing their power to adjudicate the matters brought before them. The inconsistencies between the two High Court decisions also ensure that the definitive indicators of judicial process will remain elusive. Where a minor development in the law of false imprisonment would have sufficed, the Court chose to perpetuate confusion.

⁴² It already extends to orders made by judicial officers beyond jurisdiction: *Von Arnim v Federal Republic of Germany [No 2]* [2005] FCA 662 (3 June 2005) [6].

*Andrew Lynch**

***JUDICIALIZATION OF POLITICS: THE INTERPLAY OF
INSTITUTIONAL STRUCTURE, LEGAL DOCTRINE AND
POLITICS ON THE HIGH COURT OF AUSTRALIA***

**By Reginald S Sheehan, Rebecca D Wood and Kirk A Randazzo
Carolina Academic Press, 2012**

216 pp

ISBN 978 1 61163 207 1

The politics of the United States of America is singularly alluring to external observers. An inextricable part of this vibrant and frequently polarised landscape is the Constitution of that country and the Supreme Court which interprets it. The Court, of course, has its own ‘politics’ and has been home to a cavalcade of colourful personalities since its establishment over two centuries ago. Certainly in Australia at least, one suspects that many political scientists and constitutional lawyers view the Supreme Court with fascination laced with a measure of wistful envy — the latter arising from the sense that our own High Court of Australia is a dryly uninteresting institution by comparison.

But while Australia’s lack of a bill of rights has undoubtedly meant that the High Court’s decisions have generally failed to resonate in the public’s consciousness in a way similar to those of its American counterpart, it is a mistake to lose sight of the fact that the Court has played a central role in some major Australian political dramas: most notably through its decisions over bank nationalisation, the banning of the Communist Party, the protection of Tasmania’s Franklin river and the waterfront dispute of the late 1990s. To that list, we must add, of course, the Court’s decision in *Mabo v Queensland [No 2]*.¹ *Mabo* was, by any standard, a landmark case, the impact of which was both immediate and seismic. Quite aside from the remarkable and invigorating reasoning employed by the majority judgments, the Court’s decision was the direct catalyst for a heated political and community debate which culminated in the introduction of the *Native Title Act 1993* (Cth) by the Keating government.

The story of the *Mabo* case — its origins and aftermath — is used to open and conclude a new book by three American authors whose aim is to ‘examine how institutional and legal change impacts the evolution of a high court and its politics’ and to ‘provide the first comprehensive examination of the business of the High

* Professor, Gilbert + Tobin Centre of Public Law, University of New South Wales.

¹ (1992) 175 CLR 1 (*Mabo*).

Court'.² It is not without precedent for scholars from the United States or elsewhere³ to turn their gaze upon Australia's highest court — but it certainly seems to run counter to the natural order of things and remains sufficiently rare as to justify serious attention when it occurs.

The full title of this slim volume is *Judicialization of Politics: The Interplay of Institutional Structure, Legal Doctrine, and Politics on the High Court of Australia*. The book's publisher, Carolina Academic Press, appears set on cornering the modern market for books about the High Court of Australia written by American political scientists — having produced Jason Pierce's *Inside the Mason Court Revolution* in 2006.⁴ That book received much attention in Australia due, in large part, to Pierce's extensive use of material gathered from oral interviews with many past and serving members of the Australian judiciary, as well as leading practitioners. Likened by one newspaper columnist to 'a transcript of bugged conversations from a sweaty judicial locker room',⁵ the candour, and indeed pungency, of some of the interview excerpts excited both speculation as to the identity of individual speakers and assertions, later rejected by Pierce,⁶ of a deeply divided judicial culture. Although an unquestionably valuable study, one problem with *Inside the Mason Court Revolution* was that several of the interviewees expressed themselves in such a florid fashion that it risked the simultaneous magnification and simplification of genuinely held disagreements about the judicial role and the proper boundaries of the Court's power.

One might have thought that the authors of *Judicialization of Politics*, in their eschewal of any interview material to augment their reliance on highly elaborate empirical data, would run no similar danger. After all, in preference to the 'sweaty judicial locker room' they have headed to the computer laboratory. But whether setting the scene for the statistical results they present or drawing conclusions from the same, the authors frequently rely upon generalisations of concerning breadth (referenced, it must be said, to some occasionally rather curious

² Reginald S Sheehan, Rebecca D Wood and Kirk A Randazzo, *Judicialization of Politics: The Interplay of Institutional Structure, Legal Doctrine, and Politics on the High Court of Australia* (Carolina Academic Press, 2012) 7.

³ Other United States studies are referred to later in this review, but for an example of very highly regarded scholarship from Canada on the *Mabo* case itself see Peter H Russell, *Recognising Aboriginal Title: The Mabo Case and Indigenous Resistance to English-Settler Colonialism* (University of Toronto Press, 2005).

⁴ Jason Pierce, *Inside the Mason Court Revolution: The High Court of Australia Transformed* (Carolina Academic Press, 2006).

⁵ Janet Albrechtsen, 'Inside Judges' Secret World', *The Australian* (Sydney), 14 July 2007.

⁶ Jason Pierce, 'Reactions to the Mason Court: What are Australia's Judges Thinking?' (Paper presented at the Gilbert + Tobin Constitutional Law Conference, Sydney, 8 February 2008) <http://www.gtcentre.unsw.edu.au/sites/gtcentre.unsw.edu.au/files/mdocs/351_JasonPierce.pdf>.

sources).⁷ This is most apparent in their truncated and fairly lifeless account of the history and development of the High Court and its individual judges over the first two substantive chapters. So, to give just one example, reservations might be expressed about this sentence on page 12:

After Prime Minister Keating's government attempted to implement the Court's [*Mabo*] decision through the *Native Title Act* of 1993, his Australian Labor Party suffered electoral defeat at the hands of Howard's Liberal-National Coalition in 1996.

For one thing, 'attempted to implement' is a poor description of the relationship between the *Mabo* decision and the legislative response it provoked, quite aside from hardly doing justice to the achievement of the *Native Title Act*. The sentence also suggests a causal link between that legislation and the defeat of the government three years later when of course the picture is very much more complicated.

Superficiality in describing the complex historical and political significance of an evolving multimember institution such as a final constitutional court must be an occupational hazard for those working in another jurisdiction. Certainly it is not a task for the faint-hearted armed only with theories devised in another setting and a head for numbers. The Pierce book stood up not merely because its author had, I suspect, a more nuanced understanding of the phenomenon he was studying than do Sheehan, Wood and Randazzo but also, of course, much of the book was gleaned directly from the chambers of Australian judges and advocates. Without the benefit of that sort of extensive exposure to the topic, it is little wonder that *Judicialization of Politics* frequently reflects the limitations of its authors' grasp of Australian law and politics. Sometimes this leads to simple slip-ups such as confusing the cousins (Prime Minister) Robert and (Justice) Douglas Menzies; sometimes it produces understandable if still mildly offbeat remarks such as referring to a High Court judge having being knighted by the 'Queen of England'; and other times, often when they are attempting to make an important point, it is just downright strange. How else are we to view the blunt observation made of the ALP that 'their main political goals remain outside the scope of the constitutional framework'?⁸

These and other clangers in the commentary might conceivably be taken in our stride if they did not undermine the reader's confidence in the utility of the book's many and varied empirical findings. While lots of the tables and graphs presented throughout the book are illuminating on particular points, a great many of them are hindered by a lack of clarity about the categories employed to arrange the data — and doubt about the way in which the primary material of the Court's decisions might have been handled.

⁷ For example, discussion of Sir Owen Dixon's views on Sir John Latham's judgment in *Australian Communist Party v Commonwealth* (1951) 83 CLR 1 is referenced to the entry on Dixon in the *Australian Dictionary of Biography* and the identification of distinctive features of the High Court under Chief Justice Mason are referenced to newspaper opinion-editorials by Janet Albrechtsen.

⁸ Sheehan, Wood and Randazzo, above n 2, 135.

For the empirical testing of their various hypotheses, the authors of *Judicialization of Politics* drew on the National Science Foundation's High Courts Judicial Database — an American resource of which, to the best of my knowledge, no Australian researcher of any discipline has ever made use. This may say more about the modest attention given to the study of 'judicial politics' in this country than anything else. Only a handful of political scientists have written on the Court and those who have done so have taken a rather different approach to that on display in this book. To consider a non-exhaustive but prominent sample, the work of Brian Galligan,⁹ Haig Patapan¹⁰ and Katharine Gelber¹¹ is largely free of talk of 'dummy variables' and 'regression analysis'. This perhaps explains why the work of those authors has been effective in crossing the disciplinary divide and being influential upon legal scholars. Tony Blackshield's famous jurimetrics studies on the Court over the 1970s were strongly influenced by United States jurimetrics research,¹² and followed on from Glendon Schubert's studies of the High Court in the late 1960s,¹³ but Blackshield determined the classification of primary material himself. Likewise, the hugely prolific Russell Smyth, whom the authors correctly attribute with having 'almost single-handedly reinvigorated the empirical study of judicial behaviour in Australia' in the last decade,¹⁴ does not use this database but independently manages the data he presents.¹⁵

⁹ Brian Galligan, *Politics of the High Court: A Study of the Judicial Branch of Government in Australia* (University of Queensland Press, 1987).

¹⁰ Haig Patapan, *Judging Democracy: The New Politics of the High Court of Australia* (Cambridge University Press, 2000).

¹¹ Katharine Gelber, 'High Court Review 2004: Limits on the Judicial Protection of Rights' (2005) 40 *Australian Journal of Political Science* 307; Katharine Gelber, 'High Court Review 2005: The Manifestation of Separation of Powers in Australia' (2006) 41 *Australian Journal of Political Science* 437; Anika Gauja and Katharine Gelber, 'High Court Review 2010: The Resurgence of Rights?' (2011) 46 *Australian Journal of Political Science* 683.

¹² A R Blackshield, 'Quantitative Analysis: The High Court of Australia, 1964–1969' (1972) 3 *Lawasia* 1; and A R Blackshield, 'X/Y/Z/N Scales: The High Court of Australia, 1972–1976' in Roman Tomasic (ed), *Understanding Lawyers: Perspectives on the Legal Profession in Australia* (Law Foundation of New South Wales, 1978) 133.

¹³ Glendon Schubert, 'Political Ideology on the High Court' (1968) 3 *Politics* 21; Glendon Schubert, 'Judicial Attitudes and Policy-Making in the Dixon Court' (1969) 7 *Osgoode Hall Law Journal* 1.

¹⁴ Sheehan, Wood and Randazzo, above n 2, 136.

¹⁵ See, by way of a small sample, Russell Smyth, "'Some are More Equal than Others' — An Empirical Investigation Into the Voting Behaviour of the Mason Court' (1999) 6 *Canberra Law Review* 193; Russell Smyth, 'Judicial Interaction on the Latham Court: A Quantitative Study of Voting Patterns on the High Court 1935–1950' (2001) 47 *Australian Journal of Politics and History* 330; Russell Smyth, 'Historical Consensual Norms in the High Court' (2002) 37 *Australian Journal of Political Science* 255; Russell Smyth, 'Explaining Historical Dissent Rates in the High Court of Australia' (2003) 41 *Commonwealth & Comparative Politics* 83.

I am not questioning the integrity of the material gathered in the National Science Foundation's High Courts Database, but on the evidence provided by *Judicialization of Politics*, I think it is fair to say that the way in which High Court decisions have been 'coded' in that resource is problematic — at least it is intended to be enlightening to those who regularly follow the Court's work. For example, in Chapter 4, the Court's decisions are broken down between 1970–2003 into four 'general issue areas': 'Criminal Law', 'Civil Liberties', 'Administrative Law' and 'Economic and Torts'. 'Administrative Law', it soon becomes clear, refers to a category more expansively badged as 'Government Regulation'. This comprises (in order of frequency): 'Taxation', 'Health, Safety & Environment', 'Unions & Labour Relations', 'Agriculture & Land Reform', 'Immigration & Citizenship' and 'Federalism'. Precisely what kind of case merits the classification of 'Agriculture & Land Reform' is anyone's guess — and the idea that in the relevant period 9.4 per cent of the High Court's 'Government Regulation' decisions were of this ilk while just 6.0 per cent concerned 'Federalism' begs several additional questions. Here is just one: how have the authors dealt with the fact that many disputes across all these various categories possess a 'federal' dimension? The suggestion from the data set is that insulation of cases to these discrete categories is not just possible, but also meaningful. I think neither is so. One need only consider s 109 of the *Constitution* to envisage how a case might concern overlapping Commonwealth and State regulation in the space of health, safety and the environment. How is such a case categorised?

'Federalism' is used in the tables of later chapters as a category apart from 'Criminal', 'Economic' and 'Tort', but from a public law perspective, this is a bizarrely narrow label to affix to what I can only presume is meant to be constitutional cases more generally. But perhaps not? It may well be that the authors have captured much of the High Court's ch III jurisprudence under the category of 'Criminal' cases. Who knows? They may have felt it too much to require the reader to wade through elaborate detail about the methodological choices made in preparing and presenting their data, but frankly I think that omission was a grave mistake. The organisational categories are insufficiently explained and raise questions about the results presented by the authors to such a degree that I doubt any Australian researcher would feel much confidence in using them. That hesitancy may vary depending on the field in which one works — but as a public lawyer, my view is that the results presented in this book which depend on the classification of cases by topic are unable to be cited to any meaningful end.

Not all of the data in *Judicialization of Politics* is of this kind. But then the authors' suppositions or conclusions present further difficulties. For example, they confirm that, as already established by Matthew Groves and Russell Smyth,¹⁶ the length of judicial opinions is increasing. They suggest, not implausibly, that this follows from the removal in 1984, with the introduction of the special leave requirement, of

¹⁶ Matthew Groves and Russell Smyth, 'A Century of Judicial Style: Changing Patterns in Judgment Writing on the High Court 1903–2001' (2004) 32 *Federal Law Review* 255.

the need for the Court to devote resources to ‘simple cases’ coming to it from the mandatory docket.¹⁷ I suspect there is much in that, but I am also persuaded by the view of Justice Ruth Bader Ginsburg, observing the same phenomenon in respect of the United States Supreme Court, that the ease of electronic retrieval of case authorities and the greater number of judicial associates have been significant drivers of this trend.¹⁸ There is no recognition of these more mundane explanations in the authors’ discussion. Nor do they really dwell on the significance of the unwavering climb in judgment length over the period under study — well after the end of the era marked by Sir Anthony Mason’s Chief Justiceship.

This last point reflects an almost unfathomable deficiency in the book which is the thinness with which it discusses the Court under Chief Justice Murray Gleeson. In part this is a matter of substance — although the authors acknowledge the return to a rhetoric of ‘legalism’, they also suggest that this could not quite mask or effectuate a retreat from the policy orientation of the Court that occurred under Chief Justice Mason. Indeed, they go so far as to make the startlingly unqualified claim that although the Mason Court may have ‘judicialized’ politics, ‘the Court has stayed in the centre of policy-making in Australia’.¹⁹ With respect, there is great complexity here that is being glossed over. But the limitation is also simply the result of the authors going no further in their study than the year 2003. I am not sure how American academia would describe this nor whether that kind of time lag is acceptable when publishing in that jurisdiction, but I would submit that by local standards we would say the *Judicialization of Politics* was already ‘out of date’ when it appeared — three years into the tenure of Chief Justice Robert French. It is not at all clear why the second, arguably more dynamic, period of the Gleeson Court is excluded from this study — and indeed, if the authors are really interested in institutional transformation, would they not want to give readers as much of a sense of the developments after the Mason era as possible? To merely offer a few curt observations about the Gleeson Court’s ‘conservatism’ is inexcusably one-dimensional. This book appeared 12 years after Haig Patapan’s *Judging Democracy* and despite the benefit of eight more years of the High Court under Chief Justice Gleeson upon which to draw, it barely advances our understanding of that era any further. The deficiency is particularly glaring when one contrasts the substantial attention given to the Gleeson period, still then on foot, by the Pierce book.

The final chapter is undoubtedly the book’s most ambitious and original. It aims to reveal decision-making trends in the Court as a matter of substance. It does so through a series of tables dauntingly titled ‘Random Intercept Hierarchical Logit Model of Voting Behaviour’ and which attempt to crunch figures based upon the political party that appoints a judge, his or her age and also state of origin (or, to be precise New South Wales versus the rest of the country) while controlling variables such as whether the case involves torts (‘tort law has been used to further civil

¹⁷ Sheehan, Wood and Randazzo, above n 2, 116.

¹⁸ Ruth Bader Ginsburg, ‘Remarks on Writing Separately’ (1990) 65 *Washington Law Review* 133, 148–9.

¹⁹ Sheehan, Wood and Randazzo, above n 2, 61.

liberties’),²⁰ the presence of government regulation and if the appeal is allowed or dismissed. The authors do express their work with some caveats but to be honest one either goes for this sort of thing or one does not. The exercise defeated me once I reached the point where the authors explained that their weighting of the variable of statehood was based around the idea that judges drawn from New South Wales ‘are expected to cast more pro-defendant votes’ in criminal cases because they will have a “‘big city” approach [which] may be associated with more lenient views on crime and punishment than the more conservative areas’.²¹ Aside from querying whether judges appointed from New South Wales necessarily identify as Sydneysiders if they have built a practice in the city but have been raised elsewhere in the State, the hypothesis concerning attitudes to crime was, at least to this reviewer, presented unconvincingly (Canadian studies are referenced but none in support of the ‘big city–pro-defendant’ connection specifically). So interrelated are the elements of the statistical findings at this point of the book that once scepticism takes hold in relation to a part, it taints the whole.

Overall, it is hard to escape the feeling that *Judicialization of Politics* will struggle to make much impact on an Australian audience — in either law or political science. The era it reviews in the Court’s history has been the subject of a sizeable number of books and articles by now and the decade since the authors drew a line under their work — the end of the Gleeson Court and the start of Chief Justice French’s tenure — has provided rich new material for analysis. The authors’ claim that *Mabo* was the signature moment in which the Court signalled a lasting ‘change to policy-oriented activism’²² is a position that would have been bold, albeit appreciable, in the mid-1990s. It is a contribution that is, at this remove, hard to press so emphatically. *Mabo* was 22 years ago and the authors look no further than the 11 years immediately following it being handed down. There is much evidence from the period they examine to counter their baldly stated claim, but much more water has run under the bridge since then.

Quite apart from timeliness, this book suffers from difficulties against which comparativists should remain ever vigilant. Not only is its attempt to give a potted history of the High Court and its judges unsatisfactorily sketchy, but the way in which its empirical data has been developed reflects usage of categories and assumptions that do not have much obvious purchase in Australia. That is, ultimately, a great shame — for there is much industry on display here and a genuine attempt to enlighten. That so much effort should be expended for so little gain, is a cause for regret.

²⁰ Ibid 154.

²¹ Ibid 153.

²² Ibid 109.

*Sheelagh McCracken**

**AUSTRALIAN PERSONAL PROPERTY
SECURITIES LAW**

by Anthony Duggan and David Brown
LexisNexis Butterworths, 2012
xxxvii + 371 pp
ISBN 978 0 409330 32 8

Almost two years ago, the long anticipated *Personal Property Securities Act 2009* (Cth) ('PPSA') finally became fully operative. With its expanded functional definition of what is now to be regarded as a security interest over personal property, its new process for making security interests effective and its changed priority and enforcement regimes, the *PPSA* radically alters the landscape of commercial law in Australia. The publication of this new book is thus warmly welcomed, not only for its timeliness but also for its extensive coverage and sophisticated analysis of the legislation.

The book opens with a useful introductory discussion of the nature of secured credit and the historical development of secured transactions law in Australia. By way of structure, it then adopts a highly practical transactional approach; with some further 14 chapters systematically examining the issues typically involved in analysing a secured transaction. As the authors observe in the Preface:

Unlike the ordinary run of statutes, it is not possible to read the PPSA from cover to cover and come away with a working knowledge of what it is about. The PPSA has its own internal logic which requires mastery before its secrets can be unlocked.¹

The early chapters (Chapters 2–6) move logically from consideration of the property over which a security interest can be taken, through the wide-ranging scope of the security interest, to the critical concepts of attachment, enforceability against third parties and perfection. These chapters focus clearly and carefully on the statutory criteria that must be satisfied if a secured party is to obtain the greatest extent of protection afforded by the *PPSA* against claims of third parties, with particular attention paid in a separate chapter (Chapter 6) to the registration process. Three subsequent chapters (Chapters 7–9) are devoted to thoroughly exploring the position of the secured party vis-à-vis other competing secured parties under the *PPSA*'s elaborate framework of priority rules. Dealings with the collateral by the grantor

* Professor, Sydney Law School, The University of Sydney.

¹ Anthony Duggan and David Brown, *Australian Personal Property Securities Law* (Lexis Nexis Butterworths, 2012) xviii.

and the ability of a buyer or lessee from the grantor to take the collateral free of a secured party's security interest are examined in close detail in Chapter 10, with Chapter 11 explaining how and when the secured party's security interest can attach to the proceeds of such dealings. The secured party's ability to realise the collateral on a default is the focus of Chapter 12. The remaining substantive chapters cover the impact of insolvency (Chapter 13), conflict of laws issues (Chapter 14) and transitional matters (Chapter 15). The book concludes by setting the Australian legislation in the context of international developments (Chapter 16) and by drawing up a very informative table (Chapter 17) setting out policy issues as well as drafting problems that the authors have identified in the *PPSA*, bearing in mind the review of the operation of the *PPSA* required to be completed within three years.²

The book has many strengths, not least of which is the clear and succinct style in which it is written. Importantly, it does not shy away from a detailed examination of the many technical rules governing security transactions that must be absorbed if the *PPSA*'s operation is to be properly understood. Indeed, the book succeeds admirably in making such rules readily comprehensible through, on the one hand, its explanation of the underlying policy which has shaped the particular rule and, on the other, the provision of carefully worked examples by which it illustrates the manner in which the rule may apply. In such a comprehensive work addressing many complex issues, it would be surprising if there were not on occasions specific propositions or conclusions with which one disagreed. Any such disagreement cannot, however, detract from the obvious breadth and depth of knowledge and scholarship that informs the analysis and discussion throughout the book. In his Foreword, Emeritus Professor Sir Roy Goode describes the work as one of 'astonishing erudition'.³

The book is without doubt a very important addition to the rapidly increasing literature on the Australian legislation and it is highly recommended for all interested in secured transactions. Legal practitioners should find that it provides much 'food for thought', especially in its reasoned views on the practical matters relating to registration, including the concern expressed over emerging local practice with respect to formulating collateral descriptions.⁴ Those involved in consumer lending transactions will value the various discussions of the *PPSA*'s interrelationship with the National Credit Code. As a teaching tool, the book should become a standard text. Undergraduate students who are not yet experienced in commercial practice will welcome in particular the helpful descriptions of the commercial context in which transactions typically arise as well as the use in the examples of familiar collateral such as motor vehicles. This reviewer has had no hesitation in prescribing the book this year for courses on secured transactions law at both undergraduate and postgraduate level.

² *Personal Property Securities Act 2009* (Cth) s 343.

³ Sir Roy Goode, 'Foreword' in Anthony Duggan and David Brown, *Australian Personal Property Securities Law* (LexisNexis Butterworths, 2012) v, v.

⁴ Duggan and Brown, above n 1, 133.

Finally, one further noteworthy feature of the book, which is of great benefit to practitioner and student alike, is its use of comparative materials. Particularly in this early stage in the *PPSA*'s development, the inclusion of analysis of case law from Canada and New Zealand as well as direct comparisons with these jurisdictions' statutory provisions is of considerable interest. Although the extent to which Australian courts will explicitly refer to the position in other jurisdictions with comparable legislation remains unclear, familiarity with these materials brings significant insights as to possible legislative and judicial approaches to common issues. Moreover, and perhaps even more importantly, such comparative materials inevitably provoke discussion of the underlying legal concepts. The authors' position on the nature of a security interest arising under the *PPSA*, seemingly based on Canadian argumentation, provides just one example. They contend that the security interest should be understood as a statutory hypothecation.⁵ Where the particular security interest arises under a general security agreement such a contention is uncontroversial. More problematic, however, is its application to a transaction in which title to property is reserved under a conditional sale arrangement or a finance lease. The concept of a unitary security interest undoubtedly requires a security interest to be subject to the same statutory rules, irrespective of how it arises; whether or not such a concept necessarily under Australian law requires each interest to be characterised in the same manner is less clear.

Such conceptual discussions, and others of a similar ilk raised by the authors, are not merely of academic interest. The *PPSA* is certainly one of the most critical pieces of legislation to be introduced into Australian commercial law for many decades. It is important, looking ahead, to ensure that its general concepts are fully considered and discussed if the legislation is to be successfully integrated into the Australian commercial law framework.

⁵ See, eg, Duggan and Brown, above n 1, 44–5.

SUBMISSION OF MANUSCRIPTS

In preparing manuscripts for submission, authors should be guided by the following points:

1. Submissions must be made via email to the Editors or Book Review Editor, or to the Publications Officer <panita.hirunboot@adelaide.edu.au>, or via ExpressO <<http://law.bepress.com/expresso/>> or Scholastica <<http://scholasticahq.com>>.
2. Authors are expected to check the accuracy of all references in their manuscript before submission. It is not always possible to submit proofs for correction.
3. Biographical details should be starred (*) and precede the footnotes. They should include the author's current employment.
4. Submissions should comply with the *Australian Guide to Legal Citation* (Melbourne University Law Review Association, 3rd ed, 2010) <<http://law.unimelb.edu.au/mulr/aglc>>.
5. An abstract of between 150 and 200 words should also be included with submissions (excluding case notes and book reviews).
6. Any figures in manuscripts that are of too low a resolution to produce a suitable print quality will be re-drawn by the *Adelaide Law Review*'s typesetters at the author's cost.
7. If the submission is accepted by the *Adelaide Law Review*, it will be published in hard copy and electronically.
8. Authors must sign an Author Agreement (available at <<http://law.adelaide.edu.au/review/submissions/>>) prior to the publication of their submission. The Editors prefer that a signed Author Agreement be included at the time of submission.

