# TABLE OF CONTENTS

**PREFACE**  
vii

**Rationality is dead! Long live rationality! Saving rational basis review**  
Michael Bishop  
1

**The content and justification of rationality review**  
Alistair Price  
37

**Taking diversity seriously: Religious associations and work-related discrimination**  
Patrick Lenta  
75

**On the fragility of associational life:**  
A constitutive liberal’s response to Patrick Lenta  
Stu Woolman  
111

**Migration, street democracy and expatriate voting rights**  
Wessel le Roux  
141

**Constitutional patriotism or constitutional nationalism? A response to Wessel le Roux**  
Karin van Marle  
167

**Does transformative constitutionalism require the recognition of animal rights?**  
David Bilchitz  
173

**Animal rights and the interpretation of the South African Constitution**  
Thaddeus Metz  
209

**Is there a difference that makes a difference between ubuntu and dignity?**  
Drucilla Cornell  
221

**Where dignity ends and ubuntu begins: An amplification of, as well as an identification of a tension in, Drucilla Cornell’s thoughts**  
Yvonne Mokgoro and Stu Woolman  
241
PREFACE

We do things differently around here. (Or rather, we like to think we do.) Throughout the year, we have jurists and academics (local and foreign), graduate students and school learners, arrive at our offices on Constitution Hill to present papers that either make us squirm in our seats or think of ostensibly mundane matters afresh. Often enough, our guests deliver presentations on topics we have yet to consider and are, therefore, new to the South African legal landscape. The creation of a safe space to push the boundaries of legal thought remains the raison d’être for the South African Institute for Advanced Constitutional, Public, Human Rights and International Law (SAIFAC).

One of the ways that we do things differently is to replace static presentations with vibrant conversations. Throughout our colloquia and seminars, as well as in other publications (from Constitutional Conversations to the Constitutional Court Review), we have employed a dialogical model of presentation. Instead of having individual judges, academics and students present alone, we opt for formats that place individuals in conversation with one another – in a manner that is rigorous yet collaborative, spirited rather than competitive.

One new result of this methodology is this book. The colloquies that make up this work are framed in a manner that requires interlocutors to speak directly to, and not past, one another. Professor Frank Michelman of Harvard Law School (one of the authors in this volume) neatly captures the colloquium’s ethos: ‘The aim is to learn. It is aggressively to learn what there is to be learnt from puzzles [our] interlocutors pose to us, by assuming there is method in their madness and doing our best to ferret that out, using everything else we know or can guess (in part from their likeness and kinship to us) about where they are coming from.’

Hard as it may be to put our own thoughts into a coherent, compelling fashion, harder still is it to credit, and to give at least initial priority to, the claims of others who stand in apparent disagreement with our own positions. To do so takes work. Real work. And patience. The authors of this collection of colloquies are decidedly not out there playing games of ‘gotcha’. They are engaged in a collective effort to close down areas of difference (even as they take risks associated in opening up new areas of exploration). The purpose in crediting our interlocutors as we do is to sharpen the critical bite of outstanding differences. Sometimes a pair of conversants will wind up singing off the same hymn sheet. In other dialogues, presenters are forced to rethink their initial positions, and provide reassessments or clarifications thereof in footnotes that address the challenges posed by their respondents. Again: We hope that this format will help illustrate the complexity of certain topics, whilst highlighting areas of agreement and disagreement in others.

The topics are diverse but all engage matters of great import for South African constitutional law. Some authors ask us to reconsider our disciplinary bias – as lawyers – by asking questions about the relationship between theory and practice across diverse forms of life, and whether law...
is fundamentally different than other norm driven social practices (such as religion, golf or psychoanalysis). Others hold fast to the received wisdom that law is an autonomous domain in which theory and practice differ from other ways of being in the world. Other authors discuss questions of constitutional doctrine that seem deceptively simple on their face: ‘rationality review’. Both authors, in this colloquy, leave us with entirely fresh understandings of how we should read the Constitutional Court and develop the doctrine in this domain. At the same time, one cannot be faulted for thinking that far too much ink has been spilled on the desirability of ‘balancing’ in our constitutional jurisprudence. And yet we are here asked to engage with the nuanced conception of proportionality and balancing gleaned from the well-known work of Robert Alexy. One of the more hotly contested debates in our current jurisprudence concerns the place of the value of ubuntu – and its relationship to dignity and other rights in our constitutional order. This colloquy’s constructive engagement between a former Constitutional Court justice, a critical theorist and a ‘run-of-the-mill’ constitutional law scholar does not lead us down the intellectual dead end of cultural relativism and value incommensurability. Instead, it demonstrates the way in which a dominant norm in Western ethical discourse can itself be illuminated by comparison and contrast to a central feature of African thought.

Well, if dignity has its day, equality must also have its say. In a ground-breaking exchange on whether the Bill of Rights applies to non-human animals, the authors ask whether equality, and the underlying ethic that demands rejection of all forms of arbitrary discrimination, ought not as a matter of consistency be applied to the ongoing crimes committed against many of the animals that share our society. In another bracing exchange, two authors seek out the best justification for a democratic constitutional order – toleration, diversity, social capital – and the ramifications those justifications have for questions of differentiation and discrimination in religious communities.

In a particularly challenging conversation, two authors ask us to reconsider our notion of ‘citizenship’ and whether its expansion beyond a discernible territorial republic is possible, let alone desirable. Other colloquies deal with subjects that initially feel more familiar. The apparently well-worn topic of judicial deference in socio-economic rights cases creates a catalyst for two fresh looks at this area in which everyone wants to know: What social entitlements can we expect to receive when all is said and done? Our constitution identifies openness and democracy as foundational values. Two interlocutors ask how open our democracy is and whether the judiciary can play a meaningful role in maintaining the transparency and the accountability of the coordinate branches – particularly in the face of a constitutional crisis. As if genuine finality could be achieved in any of these colloquies, we are asked to think again about how much clarity the Constitutional Court owes its constituents in a democracy committed to the rule of law. The exchange in this debate (that has now spread beyond the bounds of this extended colloquy) engages the coherence of some of the Court’s judgments and the expectation of clear precedents for those of us who wish to conform our behaviour to the dictates of the basic law.
Finally, real communities and sustained conversations – unlike ideal communities and imagined interlocutors – require real resources. We would like to thank the Konrad Adenauer Stiftung for their financial support of the *Is This Seat Taken? Conversations at the Bar, the Bench & the Academy* series and colloquia – and its regular support for other similar endeavours. In January 2011, SAIFAC became a Centre of the University of Johannesburg and we are grateful for the continued support of the Faculty of Law for SAIFAC’s activities.

Discrete grants from the World Bank and *Constitutional law of South Africa* have made dissemination of these colloquies possible in this format. Many of the exchanges were previously published in 2010 in a special edition of *Southern African Public Law*. We are grateful to the editors of the *SAPL* – both for their initial collaboration and their subsequent permission to re-publish the majority of the articles that appear in this book. The *South African Journal on Human Rights* and the *South African Law Journal* have, likewise, graciously allowed for the reprint of previous exchanges here. We would like to express our deepest appreciation to Lizette Hermann, our Pretoria University Law Press publisher, and SAIFAC Researcher, Juha Tuovinen, for their hard work in seeing this book into print. Finally, no event or publication at SAIFAC would ever come to fruition without the ongoing commitment of staff members Dolores Joseph and Vusi Ncube to the smooth day-to-day functioning of the Institute.

**Professor Stu Woolman**
*University of the Witwatersrand*

**Professor David Bilchitz**
*University of Johannesburg*
*South African Institute for Advanced Constitutional, Public, Human Rights & International Law*
*The Old Fort, Constitution Hill, Braamfontein, Johannesburg, South Africa*

March 2012
CHAPTER 1

RATIONALITY IS DEAD!
LONG LIVE RATIONALITY!
SAVING RATIONAL BASIS REVIEW*

Michael Bishop**

[T]he word ‘test’ is inappropriate, at least insofar as it suggests some meaningful analytical framework to guide judicial decision making, because the rational basis test is nothing more than a Magic Eight Ball that randomly generates different answers to key constitutional questions depending on who happens to be shaking it and with what level of vigor.¹

1 Introduction

Rational basis review has been in trouble for a while.² For at least the past 35 years it has been constantly criticised for being ‘empty’,³ ‘toothless’,⁴ inconsistent⁵ and incoherent.⁶ Despite its doubtful pedigree, litigants

---

* I would like to thank Alistair Price for agreeing to reply to this paper and for the many conversations which I can only hope were as enlightening for him as they were for me. This paper was originally written for a seminar on public law at Columbia Law School run by Prof Gillian Metzger and Prof Trevor Morrison. I would like to thank them both for giving me the opportunity to write the paper and for their helpful comments on the original draft. I would also like to thank Stu Woolman, James Fowkes, the anonymous referees and all the participants at the SAIFAC seminar for their comments. All errors, of course, remain my own.

** BA (Law) LLB LLM (Pretoria) LLM (Columbia); Visiting Research Fellow, Columbia Law School; Extraordinary Lecturer, Department of Public Law, University of Pretoria.

¹ C Neily ‘No such thing: Litigating under the rational basis test’ (2005) 1 New York Journal of Law and Liberty 898 898.
² See, eg, H Linde ‘Due process of lawmaking’ (1975) 55 Nebraska LR 197; G Gunther ‘Foreword: In search of evolving doctrine on a changing court: A model for a newer equal protection’ (1972) 86 Harvard LR 1; Note ‘Legislative purpose, rationality, and equal protection’ (1972) 82 Yale LJ 123; R Bennett ‘“Mere” rationality in constitutional law: Judicial review and democratic theory’ (1979) 67 California LR 1049.
³ See, eg, n 2 above 128.
⁴ See, eg, Gunther (n 2 above) 18 - 19; S Bice ‘Rationality analysis in constitutional law’ (1980) 65 Minnesota LR 1 at 3 - 4; In re Agnew (7th Cir 1998)144 F 3d 1013 1014.
⁵ See, eg, Bennett (n 2 above) 1060; R Farrell ‘Legislative purpose and equal protection’s rationality review’ (1992) 37 Villanova LR 1 2.
continue to rely on it and the courts refuse to alter the test to answer their critics. This essay is one more in a long lineage of attempts to figure out what is wrong with rational basis review and propose a way forward.

Although I rely extensively on earlier work in the area, I have taken a different tack. Firstly, I offer a comparative perspective that looks at the law in both the United States and South Africa. The formulation of the tests and judicial attitudes toward their enforcement are virtually identical in both countries. Yet the US has a much longer history of cases and scholarship that address the problem that can teach South African courts as much from its failures as its successes. South Africa also makes an interesting comparison from the US point of view, because it illustrates how even a system without signs of the extreme excesses of US jurisprudence is in need of saving and may offer some hints on what the route to salvation might be.7

Secondly, my goal is not to identify all the inconsistencies and absurdities of rational basis theory8 or practice.9 Instead I want to show how the test fails to do what it is meant to do and to suggest that a re-think of how we justify and conceptualise the test is required. I begin in part 2 by specifying the form of the test and noting some of the major differences between US and South African jurisprudence. Part 3 examines the justifications for rational basis review and deduces what type of test those justifications envisage and what types of laws they require to be struck down. The heart of my analysis comes in part 4 where I describe how the rational basis test is incapable of uniform or routine application; its outcomes cannot be neatly deduced from a formula but are almost entirely dependent on the discretion of litigants, lawyers and, most importantly, judges. Part 5 demonstrates how the analysis in part 4 unmoors the rational basis test from its traditional justifications. Finally, in part 6, I suggest an additional justification for the test that is compatible with the

---

7 In his excellent reply, Price notes that there is an important difference between equality analysis in the two countries: many laws that are tested under rationality review in the US will be tested under the more exacting unfair discrimination standard in South Africa. A Price ‘The content and justification of rationality review’ in S Woolman & D Bilchitz (eds) Is this seat taken? Conversations at the Bar, the bench and the academy about the South African Constitution (2012) 37. This is true and perhaps explains why rational basis review has attracted much more attention and criticism in the US than in South Africa. However, there is still a significant class of cases – differentiation on grounds not listed in sec 9(3) or analogous grounds – that is subjected to rationality review in both nations. Those are the primary subject of this paper. Price also notes that the criticisms leveled at American courts’ treatment of rationality review cannot be uncritically transposed to South Africa. Again he is correct. South African courts have, thus far, been far more circumspect in their application and explication of rationality review than their American counterparts. There are places in the paper where I unfairly lump all the sins of one on both. If I could rewrite it, I would change that. However, I would argue, and I would hope this paper demonstrates, that the way the test is structured and the justifications offered in favour of it are virtually identical on both sides of the Atlantic.

8 See, eg, Note (n 2 above); Linde (n 2 above).

9 See, eg, Neily (n 1 above).
reality of rational basis review and suggest what an appropriate formulation of the test might look like.

Three small notes are in order before I begin. First, I am only concerned with the rational basis test in the context of equality claims. Much here might be applicable to the other contexts in which rationality review occurs, but I do not directly address those issues here. Second, Price has written a fantastic response to this essay in which he carefully tests my assertion that there is something wrong with the current groundwork of rationality review. I agree with much of what Price says and believe his work deepens our understanding of the role rationality review can and should play in a constitutional democracy. I am deeply indebted to him for his careful and generous engagement with my ideas. However, there are areas – some more important, some less so – in which we differ. In order to maintain the call-and-response character of our conversation, I do not react to Price in the text, although it certainly would be greatly improved if I did! Avid readers will, however, find my thoughts on some of his claims in footnotes. Third, since writing this paper, the Constitutional Court handed down four important decisions on rationality as an element of legality. I have not incorporated these cases into my discussion but refer to them because they will be of interest to readers of this article. However, both Price and I have discussed them and their relation to this paper’s thesis in other fora.

In short, I believe that the four decisions – especially when read together – confirm and support the analysis offered below. But this article should be read as stating the law at the end of 2009.

10 Price does tackle rationality in all its guises. I think much may be gained from considering all occurrences of rationality review together. However, I have two partially conflicting reasons for focusing on equality. One, I think that, ultimately, virtually any case in which rationality arises can be framed as a differentiation. There are technical reasons why the challenge will not be brought under sec 9(1) – if, eg, it is a challenge to a constitutional amendment – but the substance of the test is the same. Two, although any challenge can be brought as a differentiation claim, something happens when we invoke the right to equality that subtly changes the nature of the complaint. Unequal treatment is the starting point. Rationality is the standard we use to determine whether the unequal treatment – which we intuitively think of as somewhat unjust – is acceptable. When rationality is applied without the allegation of differentiation the flaw is the irrationality alone, not any additional underlying wrong. That distinction will affect both the justifications for, and I think the nature of, the test we use. These reasons justify looking at rationality separately in the context of equality, even if the vast majority of insights are translatable to, and from, other contexts.

11 Price (n 7 above).


2 The test

The formulation of the basic test is remarkably similar in both South African and US jurisprudence. The relevant provision in the South African Constitution is section 9(1) which provides: 'Everyone is equal before the law and has the right to equal protection and benefit of the law.' The South African Constitutional Court has uniformly phrased the test as whether the 'the differentiation bear[s] a rational connection to a legitimate government purpose'.\(^{14}\) Under American law, rationality review is based on the equal protection clauses of the Fifth and Fourteenth Amendments.\(^{15}\) Although it has been phrased in many different ways,\(^{16}\) the basic test is the same: There must be 'a rational relationship between the disparity of treatment and some legitimate governmental purpose'.\(^{17}\)

In addition to the virtually identical wording of the tests, there are other similarities between the two systems. Both stress the need not to second-guess the wisdom of legislative choices.\(^{18}\) According to the South African Court: 'The question of whether the legislation could have been tailored in a different and more acceptable way is ... irrelevant to the question of whether there is a sufficient relationship between the means chosen and the end sought.'\(^{19}\) Likewise, Justice Thomas stated: 'Equal protection is not a licence for courts to judge the wisdom, fairness, or logic of legislative choices.'\(^{20}\) Both courts also adhere to the line that only a very limited connection between means and ends is required\(^{21}\) and, as a result,

---

\(^{14}\) Harken v Lane NO 1998 1 SA 300 (CC) para 53. See also, eg, Prinsloo v Van der Linde 1997 SSA 1012 (CC) para 25; Jooste v Score Supermarkets Trading (Pty) Ltd 1999 2 SA 1 (CC) para 17; Weare v Ndebele NO 2009 1 SA 600 (CC) para 46; Van der Merwe v Road Accident Fund 2006 4 SA 230 (CC) para 42.

\(^{15}\) US Constitution amend. V ('No person shall be ... deprived of life, liberty or property without due process of law') and XIV, para 1 ('No state shall ... deny to any person within its jurisdiction the equal protection of the laws').

\(^{16}\) San Antonio Independent School District v Rodriguez 411 US 140 (1973) (the differentiation must 'bear some rational relationship to legitimate state purposes'); New Orleans v Dukes 427 US 297 303 (1976) ('our decisions presume the constitutionality of the statutory discriminations and require only that the classification challenged be rationally related to a legitimate state interest'); City of Cleburne v Cleburne Living Center Inc 473 US 432 440 (1985) ('legislation is presumed to be valid and will be sustained if the classification drawn by the statute is rationally related to a legitimate state interest'); Massachusetts Board of Retirement v Murgia 427 US 307 312 (1976) ('rationally related to furthering a legitimate state interest').


\(^{18}\) Jooste (n 14 above) para 17; Weare (n 14 above) para 60; Prinsloo (n 14 above) para 35. United States Rail Road Retirement Board v Fritz 449 US 166 197 (1980) (Brennan J dissenting); Dukes (n 16 above) 303; Heller (n 16 above) 319.

\(^{19}\) Prinsloo (n 14 above) para 35.


\(^{21}\) Prinsloo (n 14 above) para 26 ('it must be established that there is no rational relationship between the differentiation in question and the governmental purpose which is proffered to validate it'); Weare (n 14 above) para 46 ('The question is not whether the government could have achieved its purpose in a manner the court feels is better or more effective or more closely connected to that purpose. The question is whether the means the government chose are rationally connected to the purpose, as opposed to being arbitrary or capricious.') Dandridge v Williams 397 US 471 485 (1970)
both courts rarely invalidate laws on this basis.22

However, despite the similar wording of the test, there are important differences between the two countries. Firstly, while the formal phrasing of the test remains untouched, there is constant debate within the US Supreme Court about the degree of connection between means and ends that rationality requires. One line of cases, beginning with *FS Royster Guano Co v Virginia*, requires a ‘fair and substantial relation’ between means and ends.23 However, the dominant position seems to be that ‘[w]hen the classification in such a law is called in question, if any state of facts reasonably can be conceived that would sustain it, the existence of that state of facts at the time that the law was enacted must be assumed’.24 So far, there has been no such debate or controversy in South Africa where the Constitutional Court has uniformly adhered to the standard that the means need only be one possible way to achieve the end.

Secondly, US courts are not only permitted but also encouraged to imagine what purposes the law might serve.25 The position is well summarised in *Federal Communications Commission v Beach Communications Inc*:

> [T]hose attacking the rationality of the legislative classification have the burden ‘to negative every conceivable basis which might support it[,] … [B]ecause we never require a legislature to articulate its reasons for enacting a statute, it is entirely irrelevant for constitutional purposes whether the conceived reason for the challenged distinction actually motivated the legislature.26

In contrast, while the South African courts have not directly ruled on the issue, they seem to have taken the approach that the law must be tested only against the purposes that either appear from the face of the statute or

(‘If the classification has some “reasonable basis”, it does not offend the Constitution simply because the classification “is not made with mathematical nicety or because in practice it results in some inequality”’ *Lindsley v Natural Carbonic Gas Company* 220 US 6178 (1911). The problems of government are practical ones and may justify, if they do not require, rough accommodations – illogical, it may be, and unscientific. *Metropolis Theatre Company v City of Chicago* 228 US 61 (1913))’ quoted with approval in *Heller* (n 17 above) 321.)22 For a summary of the success rate of rational basis claims in the US, see Farrell ‘Successful rational basis claims in the Supreme Court from the 1971 term through *Romer v Evans*’ (1999) 32 *Indiana LR* 357. The only times the Constitutional Court has invalidated statutes are *Van der Merwe* (n 14 above) and *S v Ntuli* 1996 1 SA 1207 (CC). There is also a line of cases where courts have used the test to ensure equality in the legal process. See C Albertyn & B Goldblatt ‘Equality’ in Woolman et al (eds) *Constitutional law of South Africa* (2007) (2nd ed) para 9.3(c).

22 For a summary of the success rate of rational basis claims in the US, see Farrell ‘Successful rational basis claims in the Supreme Court from the 1971 term through *Romer v Evans*’ (1999) 32 *Indiana LR* 357. The only times the Constitutional Court has invalidated statutes are *Van der Merwe* (n 14 above) and *S v Ntuli* 1996 1 SA 1207 (CC). There is also a line of cases where courts have used the test to ensure equality in the legal process. See C Albertyn & B Goldblatt ‘Equality’ in Woolman et al (eds) *Constitutional law of South Africa* (2007) (2nd ed) para 9.3(c).

23 233 US 412 415 (1920).

24 *Lindsley* (n 21 above) 78 - 79.

25 See *Nordlinger v Hahn* 505 US 1 at 11 (1992); *Fritz* (n 18 above) 174 - 179; *Vance v Bradley* 440 US 9 111 (1979); *Dandridge* (n 21 above) 484 - 485.

that government advocates in the litigation. They have not engaged in the sort of speculation often evident in the US courts.

A third difference lies in the use of empirical evidence. The US courts have taken the attitude that 'legislative choice is not subject to courtroom fact-finding and may be based on rational speculation unsupported by evidence or empirical data', although they do not always adhere to that standard. A litigant bringing a rationality challenge cannot, therefore, rely on the fact that the law does not in fact achieve its alleged end if the legislature – or a court – could ‘rationally speculate’ that it might. South African courts have not yet taken a stand on this issue, although there are signs that they would consider empirical evidence.

Finally, US courts approach rational basis claims with a ‘strong presumption of validity’. South African courts have not adopted any presumptions in this area. Although the plaintiff must prove the case, the burden is the same as any other constitutional claim.

Although the two jurisdictions have much in common, if one takes the courts at their word, the US courts apply a more lenient test. This will be important in what follows because, while there are certain specific problems with some of the positions taken in the US, most of the analysis that follows applies equally to both regimes.

3 Justification

Having established the basic outline of the test, the next step is to inquire what justifies its existence and what ends it is meant to serve. The test is a means for courts to strike down legislation enacted through ordinary democratic processes and such decisions therefore require a sturdy justification.

---

27 There is no case in which a law has been upheld on a basis not argued by the state. In *Union of Refugee Women v Director, Private Security Industry Regulatory Authority* 2007 4 BCLR 339 (CC), the Constitutional Court declined to adopt a rationale that would have provided a much stronger justification for the government’s action than the one government relied on. The case concerned a law permitting only citizens and permanent residents to qualify as security guards. The government’s asserted purpose was an interest in ensuring the safety of citizens, allegedly furthered by the law because it was more difficult to check refugees’ previous criminal history. The more obvious purpose was simply to protect the job market for those who had a demonstrated commitment to the country. Yet the Court did not even mention that purpose because government had specifically disavowed any reliance on it.

28 *Beach Communications* (n 20 above) 315, quoted with approval in *Heller* (n 17 above) 320.


30 *Matatiele Municipality v President of the Republic of South Africa (1)* 2006 5 BCLR 622 (CC) (Sachs J concurring).

31 See, eg, *Murgia* (n 16 above) 314; *Heller* (n 17 above) 319; *Beach Communications* (n 20 above) 314 - 315.
Rationality is dead! Long live rationality! Saving rational basis review

There are two basic evils that the rational basis requirement is meant to avert: arbitrariness and prejudice. Most obviously, it is designed to ensure that whatever government's goals might be its actions have a reasonable possibility of achieving those goals. According to the South African Constitutional Court, it is a 'fundamental premise [ ] of the constitutional state … that the state is bound to function in a rational manner'.

This goal focuses on the achievement of legislative ends, not their substantive legitimacy. While there are instrumental benefits to this requirement, it is an end in itself: We expect our representatives not to differentiate between us arbitrarily. We want our law makers to debate the justifications for a law and to inform themselves about its probable effects before they act. The rational basis test is meant to ensure that they do so. Gunther, calling for courts to demand a closer connection between means and ends in rationality cases, put it this way:

A common defence of extreme judicial abdication is that the state has considered the contending considerations. Too often the only assurance that the state has thought about the issues is the judicial presumption that it has. Means scrutiny would provide greater safeguards that the presumed process corresponds to reality – and would thereby give greater content to the underlying premise for deferring to the state's resolution of the competing issues.

Secondly, the rational basis test is meant to prevent government from exercising 'naked preferences'. Government may not act solely to privilege a politically favourable group over an unfavourable group. The rational basis test requires that all laws that disadvantage a group must serve a public purpose or value. Sunstein explains:

As I read it, the function of the [equal protection] clause is to prohibit unprincipled distributions of resources and opportunities. Distributions are unprincipled when they are not an effort to serve a public value, but reflect the view that it is intrinsically desirable to treat one person better than

---

32 Prinsloo (n 14 above) para 25. See also S v Makwanyane 1995 3 SA 391 (CC) para 156 (Ackermann J concurring). ('The idea of the constitutional State presupposes a system whose operation can be rationally tested against or in terms of the law. Arbitrariness, by its very nature, is dissonant with these core concepts of our new constitutional order."

33 Gunther (n 2 above) 44 ('Means scrutiny, to the contrary, can improve the quality of the political process – without second-guessing the substantive validity of its results – by encouraging a fuller airing in the political arena of the grounds for legislative action. Examination of means in light of asserted state purposes would directly promote public consideration of the benefits assertedly sought by the proposed legislation; indirectly, it would stimulate fuller political examination, in relation to those benefits, of the costs that would be incurred if the proposed means were adopted').

34 As above.

35 See, generally, Sunstein (n 6 above); C Sunstein 'Naked preferences and the Constitution' (1984) 84 Columbia LR 1689 1692; C Sunstein 'Interest groups in American public law' (1985) 38 Stanford LR 29; F Michelman 'Politics and values or what’s really wrong with rationality review?' (1979) 13 Creighton LR 487; L Tribe American constitutional law (1988) 1451.
another. Such an understanding as this focuses on the reasons or motives that underlie classifications.36

This rationale has been explicitly endorsed in both the US37 and South Africa.38 In Justice Brennan’s words: ‘For if the constitutional conception of “equal protection of the laws” means anything, it must at the very least mean that a bare congressional desire to harm a politically unpopular group cannot constitute a legitimate governmental interest’.39

These bases for rationality review only make sense within a model of politics where ‘politics is “not the reconciling but the transcending of the different interests of the society in the search for the single common good”’.40 This understanding of politics – known as the ‘public value’ or ‘social good’ model – contrasts with the alternative ‘public choice’ model which assumes that the legislature is simply a ‘market-like arena’ in which individuals and special interest groups trade with each other through representatives to further their own private ends. There is no ‘public interest’, no identifiable ‘social good’; there are only bargains struck between those helped by legislation and those who are harmed.41

Reviewing legislation for a rational basis only makes sense in the public value model. Under the public choice regime, legislative decisions are necessarily arational: ‘the legislature simply does what it does – means and ends are merged’.42

While the dual demand for rationality and impartiality clearly require some standard, why is the rational basis standard the right one? The same interests could be achieved, for example, by a balancing inquiry. The rational basis test is justified as much by ensuring rational and impartial action as it is by deference to the legislature. As the courts routinely note,43 government regulation necessarily results in some groups receiving more benefits than others. If the courts required more than the most limited review of these run-of-the-mill differentiations, government would be brought to a standstill. In other words, ‘the Court’s perception of the evil at which the equal protection clause

36 Sunstein (n 6 above) 128.37 United States Department of Agriculture v Moreno 413 US 528 at 534 (1973); Romer v Evans 517 US 620 633 - 634 (1996).38 Prinsloo (n 14 above) para 25, citing Sunstein ‘Naked preferences’ (n 35 above).39 Moreno (n 37 above) 534.40 Sunstein ‘Naked preferences’ (n 35 above) 1691 quoting G Wood The creation of the American republic (1972) 58. See also F Michelman ‘Political markets and community self-determination: Competing judicial models of local government legitimacy’ (1978) 53 Indiana LJ 145 148 - 157; Sunstein (n 6 above); Michelman (n 35 above).41 Bice (n 4 above) 19 (footnotes omitted).42 As above. But see Michelman (n 35 above) 33 (arguing that rationality is also supported by an economic conception of the Constitution).43 In the US see, eg, Romer (n 37 above) 631; Beach Communications (n 20 above) 315; Royster Guano (n 23 above) 415. In South Africa, see Prinsloo (n 14 above) para 17 (‘If each and every differentiation made in terms of the law amounted to unequal treatment … the courts could be called upon to review the justifiability or fairness of just about the whole legislative programme and almost all executive conduct’).
is aimed is accompanied by a reluctance, rooted in separation of powers concerns, to conclude too readily that the evil is in fact taking place. The rational basis test is the means designed to invalidate only those laws that fall afoul of the first two rationales and no more. The rational basis test is meant to prevent excessive interference or obstruction of the legislative process and to prevent judges from taking substantive judgments outside the narrow confines of rooting out irrationality bias.

This leads to one of the central points of the argument: The rational basis test is often portrayed by courts as a mechanical test which will supply the same answers no matter who applies it. The very nature of the language – ‘rationality’ – implies that there is no room for disagreement; a person who reached a different outcome would not only be wrong, but irrational. But more importantly it rests on the requirement that judges not ‘second-guess’ legislative choices. If the test permitted or required judges to use their own discretion in determining the rationality of a choice, that command would be undermined as judges’ attitudes would pollute the supposedly neutral inquiry. The traditional underpinnings of rational basis review therefore demand a test that forces predictable outcomes and removes personal discretion. Judges are therefore obliged to frame the rational basis as excluding their discretion.

With an understanding of the compound justificatory framework for rational basis review, we can ask the next important question: Taken together, what types of laws are prohibited by the rational basis test? There is a simple answer. There are two types of laws that are prohibited: laws motivated by naked preferences, whether they have that effect or not, and laws that are motivated by an acceptable purpose but fail to achieve that purpose. The first type is prohibited by the ‘naked preference’ rationale, and the second by its ‘rationality’ partner.

These two classes of unacceptable laws seem to correspond directly to two distinct parts of the rational basis test. The test – the law must bear a rational connection to a legitimate government purpose – breaks down

---

44 Sunstein (n 6 above) 142.
45 See n 18 above.
46 Price criticises this passage for drawing conclusions that are too ‘stark’. He alleges that I have doomed rationality review to become ‘incoherent by sheer definitional stipulation’ and that the grant of limited discretion is compatible with the demands of institutional comity in Woolman & Bilchitz (n 7 above) 52. I agree with him on both counts. Excluding discretion completely is clearly both impossible and unnecessary to maintain appropriate judicial deference to the political branches of government. However, the thrust of my point remains: Judges are pushed to (and do in fact) downplay – if not completely deny – the extent of their discretion, and that point about judicial rhetoric is compatible with Price’s position.
47 Price suggests that prejudice or naked preference is best understood as a ‘species of arbitrariness’ in Woolman & Bilchitz (n 7 above) 58. On the understanding that this is a purely semantic rather than a substantive point, I am happy with that alteration; the understanding is correct as Price does not challenge the idea that the rational basis test is aimed at preventing both arbitrary intentions (naked preferences) and arbitrary actions (actions that do not achieve their purpose).
into (a) a threshold requirement that the purpose asserted is ‘legitimate’; and (b) a check that the law is related to some purpose that crosses that threshold. Laws motivated by covert favouritism are prohibited by (a) and those with innocent motivations but that serve no public good are proscribed by (b).

While this may be the most obvious way of understanding the test, courts often apply the two stages of the test in a different manner. They use the means-end aspect of the test as a device to ‘flush out’ impermissible purposes.48 In Cleburne the Court rejected a string of justifications for a refusal to permit the establishment of a home for the mentally-handicapped and concluded that the only possible motivation for the law was prejudice against the mentally-handicapped.49 As the law was motivated by an illegitimate purpose, it was invalid. There is nothing wrong with reversing the order in which the two-part test is administered as both types of invidious laws will still be caught out. However, it is important not to confuse interpreting an action rationally with evaluating the rationality of an action:

[I]n any particular case, the interpretive and evaluative uses of rationality are mutually exclusive. If the observer seeks to evaluate an actor’s cognitive function, the observer cannot identify the actor’s ends and beliefs through interpretive rationality because the succeeding ‘evaluation’ would be an empty tautological exercise.50

4 Balancing rationality 51

In this section I contend that, in order for the rational basis test to be meaningful and to achieve its purpose, the four parts of the test need to be carefully balanced. If an extreme position is taken in any of the four parts, then the whole test is in danger of becoming meaningless and of permitting the continued existence of any law no matter how silly or evil, or, the process becomes so arduous that the business of governing would

48 See Sunstein (n 6 above) 131; Farrell (n 5 above) 37; L Simon ‘Racially prejudiced governmental actions: A motivation theory of the constitutional ban against racial discrimination’ (1978) 15 San Diego LR 1041 at 1114; M Herz ‘Nearest to legitimacy: Justice White and strict rational basis scrutiny’ (2003) 74 Univ Colorado LR 1329 1366; B Swierenga ‘Still newer equal protection: Impermissible purpose review in the 1984 term’ (1986) 53 Univ Chicago LR 1454 1473.49

49 City of Cleburne (n 16 above).50

50 Bice (n 4 above) 8.

51 I want to point out here what the coming examples are meant to demonstrate. They are not meant to show, upon a thorough consideration of all the case law, that either the US Supreme Court or the South African Constitutional Court has adopted a particular position. Rather, they indicate how the rationality test is susceptible to manipulation. It may well be that, on average, a court has performed admirably, yet there remain a few cases that illustrate where the temptation for judges to over-reach exists, even if they are usually able to resist. I make this point because Price provides a compelling defence of the Constitutional Court’s record of applying the rationality test in Woolman & Bilchitz (n 7 above) 63 - 68. I am ready to concede much of what Price says, but I do not think it undermines my contention that the potential for manipulation exists.
be brought to a screeching halt. However, the effect of requiring this balance is that the test is necessarily influenced by judicial discretion and ceases to be a formal, logical exercise.

To repeat, the test is: The differentiation is rationally connected to a legitimate government purpose. The four parts of the test are:

(a) What is the differentiation or law at issue?
(b) What is the purpose it serves?
(c) Is that purpose legitimate?
(d) If so, is the differentiation connected to the purpose?

For reasons that I trust will become clear, I will consider these steps in more or less reverse order.

4.1 Legitimacy

The first element of the test is to determine which purposes are ‘legitimate’. In the traditional model of rational basis review this element operates as the door to the formal rationality check made up by the remaining three parts. However, as I explained earlier, that relationship can also be switched so that the rationality check becomes a mechanism to determine which purposes should be measured against the bar of legitimacy. Whether it functions as a threshold or a final arbiter, extreme approaches to the question of legitimacy pose perhaps the most blatant dangers to the validity of rational basis review.

The first extreme occurs when a court accepts virtually any purpose as legitimate. The decision of the Tenth Circuit in *Powers v Harris* offers a powerful illustration of this problem. The law in question required any person who wished to sell caskets to meet a number of stringent criteria including: completing sixty credit hours of specified undergraduate training; an

---

52 Price suggests a neat formula for this step: ‘A legal rule’s purpose is legitimate, under the rationality principle, if and only if it is consistent with all other legal and constitutional constraints, including the Bill of Rights and the Constitution’s “objective, normative value system”.’ This position will ‘constrain ... judges to deciding only questions of law’ in Woolman & Bilchitz (n 7 above) 47. As elegant as this structure is, I do not think it deals with the difficulties I identify below. The purpose of a law can be framed at various levels of abstraction, some of which will be legal while others will not. The structure also has this effect: It incorporates into the rationality test the requirement that the state make not only laws with constitutional effects, but always act with a constitutional purpose. The extent to which purpose alone is a valid reason to invalidate a law is controversial in the US and is largely unexplored in South Africa. Adding it to rationality review probably causes more problems than it solves. In my view, starting from the point that rational basis review is meant to prohibit ‘naked preferences’ – keeping in mind the various ways purpose can be changed – is a better way to approach the problem of legitimacy. This is perhaps an example of where my focus on equality leads me to a different viewpoint to that observed through Price’s wider lens.

53 379 F 3d 1208 (10th Cir 2004).
apprenticeship of one year; the embalming of 25 bodies; passing both a subject matter and an Oklahoma law exam; and having a fixed physical location where bodies could be embalmed, ‘funeral service merchandise’ sold and bodies viewed.\textsuperscript{54} In short, it permitted only proper funeral homes run by trained undertakers to sell caskets. Chief Circuit Judge Tacha declined to consider whether these requirements were rationally related to the purpose advanced by government in support of the regime, namely, to protect consumers from unscrupulous casket sellers.\textsuperscript{55} Instead, he found that the law really aimed at the protection of Oklahoma’s established intrastate funeral home industry from external competition and that, based on existing Supreme Court precedent,\textsuperscript{56} this was a legitimate government purpose.\textsuperscript{57} Naturally, the law achieved that purpose.

The problem here is that the Court required no additional public reason for the exercise of the naked preference for the sale of caskets by funeral homes over those who wish to sell caskets in ordinary shops, or on the web or from their garage. It may be argued that the state has an interest in ensuring that funeral homes are able to survive and that permitting unlimited competition in an important part of their business might force them to close down. The trouble with that argument is that reserving the casket business for funeral homes necessarily disadvantages those who would compete. The very essence of the rational basis test is that if the state wants to take sides in that competition it must have some purpose other than its preference for one side over the other.\textsuperscript{58}

That is exactly what the Supreme Court did in Fitzgerald.\textsuperscript{59} The challenge was to a law that imposed a higher tax on slot machines at racetracks than it did on the same machines on riverboats. Justice Breyer did not hold that the law was driven by a mere desire to privilege riverboats over racetracks. He noted that the riverboat industry was in financial trouble and needed support, and that ‘the legislators may have wanted to encourage the economic development of river communities or to promote riverboat history, say, by providing incentives for

\begin{itemize}
  \item Powers (n 53 above) 1212 - 1213.
  \item Powers (n 53 above) 1218.
  \item Powers (n 53 above) 1220 - 1221 citing Fitzgerald v Racing Association of Central Iowa 539 US 103 (2003); Dukes (n 16 above); Ferguson v Skrupa 372 US 726 (1963); and Williamson v Lee Optical of Oklahoma 348 US 483 (1955).
  \item Powers (n 53 above) 1218 - 1222.
  \item There may well be reasons that justify the decision, of which the consumer protection rationale avoided by the 10th Circuit is one. It may also be that undertakers would be unable to survive without the revenue generated from casket sales and that because it is in the public interest for undertakers to continue to operate they must keep a monopoly on casket sales. It is important to note the distinction between this purpose and the simple economic protectionism advanced by Judge Tacha. The purpose offered here rests on the public benefit provided by undertakers; the Court’s rationale would hold even if undertakers served no public purpose.
  \item n 56 above.
\end{itemize}
riverboats to remain in the state, rather than relocate to other states. These are all purposes that go beyond naked preference.

If one extreme is to permit naked preferences to count as legitimate purposes, the other extreme occurs when the courts take it upon themselves to decide whether they approve of government’s aims beyond the narrow confines of prohibiting blatant favouritism. A primary reason for adopting rationality as the appropriate level of scrutiny for the majority of classifications is to prevent courts from second-guessing every government decision. The temptation for courts to use the ‘legitimate purpose’ requirement to do just that is great, because it provides judges with a blank canvas upon which to impose their own views of what the Constitution should prohibit. At least under the rational basis test, courts in both the US and South Africa have been particularly sensitive to this risk. In Eisenstadt v Baird, the Supreme Court specifically avoided the moral question of whether banning contraception for moral reasons was a legitimate government purpose.

Despite these difficulties at the margins, it would seem that an easy solution exists: Only naked preferences are impermissible. As attractive as that solution is in theory, in practice it quickly fails because it is difficult to draw the line between favouritism and pursuit of the public good. Firstly, there may be disagreement about whether a purpose is really an exhibition of bias. In Romer v Evans, a majority of the Court struck down a state constitutional amendment that prevented the passage of any law that prevented discrimination on the ground of sexual orientation. To them this appeared an obvious example of a naked preference. However, to Justice Scalia the amendment was ‘a modest attempt by seemingly tolerant Coloradans to preserve traditional sexual mores against the efforts of a politically-powerful minority to revise those mores through use of the laws’. While we may favour one side over the other, it is not possible to say that either side is absolutely wrong.

60 Fitzgerald (n 56 above) 109. The same analysis can be conducted on all the other cases cited by the 10th Circuit in Powers (n 53 above). See Duke (n 16 above) 304 (‘The law only permitted vendors who had been operating in the French Quarter for eight years to continue to sell their wares. The Court noted that the purpose of the law was ‘to preserve the appearance and custom valued by the Quarter’s residents and attractive to tourists.’) Ferguson (n 56 above) 732 (the law only permitted lawyers to conduct debt-adjusting. Justice Black for the Court had this to say about the purpose of the statute: ‘The business of debt adjusting gives rise to a relationship of trust in which the debt adjuster will, in a situation of insolvency, be marshalling assets in the manner of a proceeding in bankruptcy. The debt adjuster’s client may need advice as to the legality of the various claims against him, remedies existing under state laws governing debtor-creditor relationships, or provisions of the Bankruptcy Act – advice which a non-lawyer cannot lawfully give him.’); and Williamson (n 56 above) 490 (the Court held that a law permitting only optometrists and ophthalmologists to fit and replace lenses was related to the purpose of promoting public health).

62 Romer (n 37 above).
63 Romer (n 37 above) 636.
Secondly, even legislation that is in fact aimed at blatant favouritism can be re-couched as a law with a legitimate purpose: ‘[A] single set of social consequences will be related to a more remote set of consequences, and at some point along the range of causality there should be some permissible goal for any conceivable statute’.\textsuperscript{64} Thus, a statute aimed at protecting optometrists can be said to protect public health,\textsuperscript{65} a law founded in xenophobia can be seen as a means to protect public safety,\textsuperscript{66} and an exemption intended to put money in farmers’ pockets can be seen as a device to promote the American economy.\textsuperscript{67} Below I will again return to this difficulty about the level of abstraction at which a purpose is phrased, but for now the point is simply that it is not enough to say that ‘legitimate’ means: ‘all those purposes that are not naked preferences’. Whether a law can be described as a naked preference depends on perspective and opinion, not logic or fact.

The solution to this may be to look at the actual motivation of legislation, that is, empirically determine why the legislators passed the law. If they passed it to protect optometrists it is invalid, but if they passed it out of concern for public health it is legitimate. Even if the test were phrased to permit this – which in the US it is not – it would seldom be useful, as determining actual purpose is virtually if not completely impossible. There is abundant literature on the problem of determining legislative motive.\textsuperscript{68} In addition to problems about the level of abstraction and multiple purposes, the biggest problem is assigning a single intent to a body with multiple members. When each of the legislators may have voted in favour of a law for a different reason, how can we possibly assign a single purpose to the law? Of course this problem exists only in the ‘public choice’ theory of politics; the ‘social good’ model on which rationality review is based avoids it and there are theories on how to deduce purposes in the value based model.\textsuperscript{69}

\begin{itemize}
  \item \textsuperscript{64} Note (n 2 above) 140.
  \item \textsuperscript{65} Williamson (n 56 above).
  \item \textsuperscript{66} Union of Refugee Women (n 27 above).
  \item \textsuperscript{67} Smith v Calvert 283 US 553 (1931).
  \item \textsuperscript{69} See Ely (n 68 above) 1226 (Ely argues for a ‘consensus theory’ which defines the purpose of a law as ‘what such laws are generally concerned with, [and] what most legislators intend to accomplish by most such laws considered in their entirety’. As Bice notes, this theory can function either as an evidentiary aid which helps to determine which purposes are more likely, or a normative aid which prevents certain purposes from being accepted (n 4 above) 32. In this context, it would function in the first fashion and would therefore be subject to judges’ own interpretations of what purpose a law is generally meant to serve. The theory also seems to fall apart in describing new, unusual or multi-pronged legislation.) Farrell (n 5 above) 4 (Farrell attempts to avoid the problem through his definition of legislative purpose: ‘purpose is not something that exists in the minds of the legislators, either individually or as
\end{itemize}
Whatever the validity of those theories, they are not based on logical deduction or empirical fact-finding, but on judicial judgment; they require judges to analyse an array of relevant factors to determine the ‘actual’ purpose behind the law.

4.2 Purpose

While the previous section focused on the difficulty of isolating illegitimate purposes, this section considers in more depth how a court can determine what the purpose of a law is because, as we will see, it has a serious impact on the means-end part of the rational basis test. I consider two main issues: the level of abstraction at which a purpose is phrased and laws that serve multiple purposes. In addition, the discussion above about the empirical difficulties of determining purpose applies equally at this stage.

4.2.1 Level of abstraction

Many scholars have noted that how generally or specifically the purpose of a law is determined will virtually predict the outcome of the rationality analysis. As Linde argues: ‘The outcome of an attack on the rationality of a law can be made to depend on whether the law is described as a means toward a somewhat remote end or as very close to an end in itself.’ It is easiest to explain this phenomenon through an example. In *Weare v Ndebele NO*, the applicant challenged a provincial law that limited gambling licences to natural persons and partnerships. The province argued that the regulation was necessary because it was more difficult to manage juristic persons and to hold them responsible for violating regulations. We can restate that purpose at infinite levels of generality. Here are six possible purposes:

(a) to promote the general welfare;
(b) to prevent economic exploitation;
(c) to regulate gambling;
(d) to make the regulation of gambling manageable and efficient;
(e) to hold all who violate gambling regulations responsible; and

a collective body. It exists, rather, as an objective concept, ‘evident in the character of the law itself,’ as an end to be achieved.’ While this may avoid the empirical problem of ascertaining what individual legislators think, it presents a further problem: How do we determine the ‘character of the law in itself’? This is what I suggest must necessarily remain an issue for judicial discretion).

See, eg, Note (n 2 above) 137 - 138; Farrell (n 5 above) 15 - 17; Bice (n 4 above) 28; Gunther (n 2 above) 47; N Wadhwani ‘Rational review, irrational results’ (2006) 84 *Texas LR* 801.

Linde (n 2 above) 212.

n 14 above.
(f) to limit licenses to natural persons and partnerships.\textsuperscript{73}

The biggest problems are with (a) and (f). Let us start with (f). When the purpose of a law is stated so specifically, it effectively reduces the means to the end. In the words of Justice Brennan:

\textit{[P]resuming purpose from result … reduces analysis to tautology. It may always be said that Congress intended to do what it in fact did. If that were the extent of our analysis, we would find every statute, no matter how arbitrary or irrational, perfectly tailored to achieve its purpose.}\textsuperscript{74}

The problem with (a) is more complicated and operates in two ways: ‘One could, with equal plausibility, conclude that there is no connection between the law and this purpose or that there is always a connection between them.’\textsuperscript{75} Firstly, the end is so far removed from the means that it is either impossible to show a connection, or meaningless to try.\textsuperscript{76} Secondly, one could argue that any conceivable law furthers some version of the general welfare and, unless courts want to get involved in defining ‘the general welfare’ – which they generally do not – they would have to accept that every law that claims this as its purpose is rational.

Although some argue that this critique only applies to (a),\textsuperscript{77} there are convincing reasons why the rationality of laws alleged to further purposes framed at the level of (b) and (c) are also impossible of meaningful measurement.\textsuperscript{78} As has been noted in a different context, ‘[t]he purpose of almost any law can be traced back down to one or another of the fundamental concerns of government: public health and safety, public peace and order, defense, revenue.’\textsuperscript{79} Once a law is brought under one of these headings, it is impossible to show that it does not achieve some understanding of that end in some way. So when courts are confronted with appeals to ‘the general welfare’, unless they are willing to take substantive positions on just what public health, or national defense demand, they cannot plausibly say that a law brought under one of those headings does not achieve the intended end.

In \textit{Union of Refugee Women}, a general appeal to public safety could justify a law preventing foreigners from working as security guards because they will be less ‘easily able to prove their trustworthiness’.\textsuperscript{80} This renders

\textsuperscript{73} This breakdown is modelled on that offered by Farrell (n 5 above) 15. See also Note (n 2 above) 137 (another example of how the levels of abstraction within a purpose can be broken down).
\textsuperscript{74} Fritz (n 18 above) 187 quoted in Farrell (n 5 above) 15.
\textsuperscript{75} Farrell (n 5 above) 16.
\textsuperscript{76} Ely (n 68 above) 1241; Bice (n 4 above) 15.
\textsuperscript{77} Ely (n 68 above) 1241 - 1248.
\textsuperscript{78} Note (n 2 above) 143 - 146 (‘Even an apparently objective goal, like the promotion of public safety, can be shown to be too vague for objective measurement’); Farrell (n 5 above) 16.
\textsuperscript{79} J Clark ‘Guidelines for the free exercise clause’ (1969) 83 \textit{Harvard LR} 327 330 quoted in Wadhwani (n 70 above) 821.
\textsuperscript{80} \textit{Union of Refugee Women} (n 27 above) para 38.
rationality review meaningless as all government has to do is pick the right goal among one of four or so basic goals and the law will be upheld.

Similarly, when the purpose is put at the level of regulating an industry or practice, every single law that prohibits some actions and permits others – or permits some people to perform an act and denies that privilege to others – vaguely related to that industry will be rationally related to the purpose. Thus, in Kotch v Board of River Port Pilot Commissioners for Port of New Orleans, the Court used Louisiana’s interest in regulating river pilotage to uphold a law that limited pilot licenses to family of existing pilots.81 It framed the purpose of the law as ‘to secure for the State and others interested the safest and most efficiently operated pilotage system practicable’.82 Limiting pilots to a family of existing pilots was held to serve this end because it promoted the ‘morale and esprit de corps’ of pilots.83 As that outcome indicates, resort to an interest in ‘regulation’ will permit any law to survive rational scrutiny.

In South Africa, the futility of setting interests at this level has been specifically recognised. In Van der Merwe v Road Accident Fund, the Constitutional Court struck down a law that prevented spouses with joint estates from claiming monetary damages from each other.84 The government tried to defend the law as a valid means to ‘regulate [the] patrimonial consequences of marriage’.85 Deputy Chief Justice Moseneke rejected this claim:

A court remains obliged to identify and examine the specific government object sought to be achieved by the impugned rule of law or provision. In other words, we are obliged to look at the specific purpose of [the law] even though the general purpose of regulating property arrangements in marriage may not in itself be open to constitutional doubt.86

To return to the example of Weare, any distinction between people would be one way of regulating gambling. If government decided to limit gambling licences to people over (or under) a certain weight, that would ‘regulate’ gambling. What is required, then, is some vision of how gambling should be regulated. While courts will be hesitant to supply that vision, it seems feasible to require government to supply a more specific purpose.

82 Kotch (n 81 above) 564.
83 Kotch (n 81 above) 563.
84 n 14 above.
85 Van der Merve (n 14 above) para 33.
86 As above. But see Jooste (n 14 above) para 17 (holding that the purpose of the law under challenge should properly be phrased as ‘a comprehensive regulation of compensation for disablement caused by occupational injuries or diseases sustained or contracted by employees in the course of their employment’. This seems to directly contradict the holding in Van der Merve that to state a purpose at the general level of regulation is impermissible).
We are left with purposes (d) and (e). Both of these purposes are set at an appropriate level because there is a potential for a court to conclude – even if it requires only the most tenuous of connections – that the means fails to serve the end. The problem is that there is no guidance for judges to know how to choose between (d) and (e) and no assurance that they will not choose one of the other purposes or something in-between. Ultimately judges have the power to manipulate the outcome of the means-ends inquiry by choosing how abstract or specifically to state the purpose. Although we can recognise flaws in reasoning and suggest guidelines, the test leaves this to the discretion of the judges.

4.2.2 Multiple purposes

Related to the problem of the level of generality at which a purpose is presented, is the difficulty that most laws serve multiple and sometimes contradictory purposes. A law that requires employees injured at work to claim limited compensation from a fund run by government and prevents them from directly suing their employer for the full damages can serve as an example. Such a law is undoubtedly intended to assure some compensation to employees should their employers be unable to pay, and to make compensation easier to recover. However, it also serves the interests of employers by protecting them from ruinous damages claims. Any proper understanding of the law must recognise both these purposes. If a court only considered the intent to benefit employees, then the provision would seem irrational because it prevents them from claiming from their employer even when the employer can afford to pay. If both purposes are considered, then the provision is a rational means to strike a balance between two competing goals.

Another scenario is where the court simply fails to consider an alternative possibility. In *Gulf, Colorado and Santa Fe Railway v Ellis* the court struck down a law that permitted successful tort claimants to reclaim attorneys’ fees only if the claim was brought against a railroad. The Court rejected a number of purposes (including ensuring that tort feasors would compensate their victims, and placing a higher duty to pay attorney’s fees on corporations) and found them all under-inclusive and therefore irrational. However, it failed to consider the alternative purpose that railroads had a history of resisting claims and that the statute was aimed only at them. This example indicates how the question of multiple purposes and levels of

---

87 See generally Farrell (n 5 above) 17 - 20; Note (n 2 above) 132 - 137; Linde (n 2 above) 209 - 210.
88 This law, the Compensation for Occupational Injuries and Diseases Act 130 of 1993, was challenged in *Jooste* (n 14 above).
89 This was the conclusion reached by the High Court. *Jooste v Score Supermarket Trading (Pty) Ltd* 1998 9 BCLR 1106 (E). The decision was overturned by the Constitutional Court which recognised the multiple interests at play. *Jooste* (n 14 above) para 17.
90 165 US 150 (1897). This example is borrowed from Note (n 2 above) 133.
91 Note (n 2 above) 133.
abstraction may merge into each other. Was the problem in *Ellis* that the Court was considering the purpose too abstractly, or that it was ignoring an alternative purpose? Either way, the conclusion is the same: courts can manipulate the outcome by choosing which purposes they recognise and how they phrase those purposes.

### 4.3 Level of connection

The most obvious point at which the rationality test can be altered is the degree of connection between means and ends. These are some of the possibilities:

Must proponents of the law demonstrate that the classification advances the legislative purpose to some extent? That a rational person would so believe? That a rational person would so believe if the facts were as the legislature supposed them to be? Or must the proponents demonstrate that the various trade-offs outlined between the costs of narrower and broader classifications are ‘rational’?

At the deferential end of the spectrum, the law only needs to potentially further its purpose in some minor way. It may also have many other effects, so long as it also has the potential to further an end. Another way of putting this is that the law may be both under-inclusive – exclude people that should be included – and over-inclusive – include people who should be excluded – as long as some of the people that are included should be included. At the spectrum’s strictest end, the law must effectively achieve the purpose in the most narrowly-tailored way possible. The strict standard is clearly too demanding and would involve too much judicial interference in government. It would also destroy any distinction between the different levels of scrutiny employed both in South Africa and the US, where stricter scrutiny is reserved for defined grounds of differentiation.

Rhetorically, at least, both US and South African courts can be situated on the most lax end of the spectrum. This extremely deferential approach has the potential to let through laws that have massive impacts on groups without proof or even an abstract likelihood of any public benefit: A mere suspicion of

---

92 Price defends the reputation of the Constitutional Court by noting how it has applied a consistent and appropriately deferential standard here in Woolman & Bilchitz (n 7 above) 65. Firstly, as my more critical evaluation of some of the cases suggests, I do not believe the level of deference applied by the Court thus far is too high. Secondly, I repeat that even if Price’s assessment that the Court has so far acted admirably is correct, that does not undermine my contention that the rhetorical form of rational basis review would permit the Court to act with less wisdom and restraint. The decisions of the US Supreme Court certainly bear out that possibility.

93 See especially Gunther (n 2 above) (where it is argued that the rational basis test should be changed to require a direct connection between means and ends and little or no scrutiny of the legitimacy of government purposes).


95 See part 2 above.
minimal advancement of a goal will do. Railway Express Agency v People of State of New York is a classic illustration of this danger.96 The Court upheld a New York law that banned advertising on vehicles except if the advertising was for the vehicle owner’s business. The differentiation was acceptable because ‘[t]he local authorities may well have concluded that those who advertised their own wares on their trucks do not present the same traffic problem in view of the nature or extent of the advertising which they use. It would take a degree of omniscience which we lack to say that such is not the case.’97 The outcome in Railway Express was not necessarily wrong; Linde has argued that the law could have been justified by considerations of equity.98 The problem is that allowing such a fragile link ‘is not judicial review but dismissal of a claim of review’;99 or, as Justice Marshall has written, the test, ‘when applied as articulated, leaves little doubt about the outcome; the challenged legislation is always upheld’.100

Perhaps driven by this recognition, the Supreme Court has occasionally applied a slightly higher standard of review that requires more than a minimal link.101

Cleburne is a paradigmatic example.102 The Court rejected a range of justifications for refusing to permit a building to be used as a home for the mentally disabled including that the home was located on a flood plain, and concerns about legal responsibility and congestion. The Court concluded that all these goals, while legitimate, applied to all people, not only to the mentally retarded, and were therefore irrational. Cleburne invalidates a law for being under-inclusive and therefore clearly sets a higher bar than traditional rationality review. Applying these rationales only to the mentally retarded would still partially further the goal and therefore satisfy basic rationality scrutiny.

The South African Court too has been guilty of departing from the minimal standard it seems to have endorsed in other cases on the two occasions it has found laws irrational. In Van der Merwe, the Court found that the distinction drawn by the statute between monetary damages and damages for pain and suffering was not ‘reliable’ and was ‘tenuous at best’.103 An honest application of the minimal standard would have upheld the law on those grounds.

96 336 US 106 (1949). See also Minnesota v Clover Leaf Creamery Co 449 US 456 (1981) (upheld a law that banned the use of plastic milk cartons but permitted paperboard cartons despite empirical evidence that plastic posed no greater environmental danger than the paperboard. The Court concluded, based on the evidence before the legislature, that as long as the question was debatable, the law had to be upheld.)
97  Railway Express (n 96 above) 110.
98 Linde (n 2 above) 210. See also Fritz (n 18 above) 178 (recognised the legitimacy of claims to equity).
99 Linde (n 2 above) 210.
100 Murgia (n 16 above) 319.
101 See, eg, Allegheny Pittsburgh Coal v Webster County 488 US 336 (1989); Moreno (n 37 above).
102 Cleburne (n 16 above).
103 Van der Merwe (n 14 above) para 57.
The departure is even more obvious in \textit{S v Ntuli}.\textsuperscript{104} Didcott J found irrational a law that required unrepresented prisoners to get a judge’s certificate before appealing their conviction while imposing no similar requirement on represented prisoners or convicted people not in jail. The purpose of the provision was to prevent appeals that were ‘lodged frivolously with a view to no gain but the opportunity for an excursion to court and some temporary relief from the tedium of imprisonment’.\textsuperscript{105} Without much analysis, Justice Didcott concluded that the scheme violated the guarantee of equal protection of the laws.\textsuperscript{106} But the minimal rationality standard would have upheld this law easily as, although some appeals by unrepresented prisoners would have been genuine and some appeals by the other classes of citizens would have been frivolous, the law would have partially furthered its end by preventing some frivolous appeals.

The results in \textit{Cleburne}, \textit{Van der Merwe} and \textit{Ntuli} can only be explained if courts adopt one, or some, combination of the following more demanding tests: (a) require the means to substantially further the end; (b) require a reason for why the state has focused only on one group and not others to achieve the end; or (c) balance the achievement of the end against the harm it causes. These methods – or similar formulations – not only explain existing precedent but also provide standards for future application. The problem is that they necessarily require courts to adjudicate on the extent to which a law must further an end or the wisdom of a legislative choice to single out a particular group. While they can still be applied with deference, they cannot be applied without discretion.

4.4 The differentiation

The discussion up to now has covered the well-trodden terrain of the difficulty in determining what purposes are legitimate, defining purposes and setting an appropriate level of scrutiny. But the rational basis test can be made to vacillate between competing goals in a fourth way: The specificity with which the law being challenged is defined. In essence, this reflects the insights about the degree of abstraction of purpose and the simultaneous serving of multiple purposes and applies them to the way the law is understood. There are two basic ways in which the law can be defined in order to induce a different result: it can be considered outside the context of other statutory provisions; only the general provision is considered.

In \textit{Union of Refugee Women},\textsuperscript{107} the majority of the Constitutional Court upheld a section of a statute\textsuperscript{108} that permitted only citizens and permanent

\textsuperscript{104} n 22 above.
\textsuperscript{105} \textit{S v Ntuli} (n 22 above) para 22.
\textsuperscript{106} \textit{S v Ntuli} (n 22 above) para 20.
\textsuperscript{107} \textit{Union of Refugee Women} (n 27 above).
\textsuperscript{108} Private Security Industry Regulation Act 56 of 2001 sec 23(1)(a).
residents to register as security guards. The majority held that this provision was rationally related to the purpose of ensuring public safety.\textsuperscript{109} It stressed that this was not based on any xenophobic tendencies:

It is not that the Authority does not trust refugees. Rather, it requires everyone to prove his/her trustworthiness. The reality is that citizens and permanent residents will be more easily able to prove their trustworthiness in terms of the Security Act.\textsuperscript{110}

Accepting the level at which the purpose is presented\textsuperscript{111} and that a minimal degree of connection was required, there is still a problem with this analysis noted by Justices Mokgoro and O'Regan in dissent:\textsuperscript{112} The statute also specifically required applicants to prove that they had not been convicted of an offence in the past 10 years. Refugees who are unable to comply with that provision will in any event be barred from being security guards and sidelining refugees who can produce proof merely because of their status as refugees will do nothing to ensure the trustworthiness of security guards.\textsuperscript{113} Only because the majority limited its challenge to the single section, instead of seeing it in the context of the statute as a whole could it find the statute rational.

\textit{United States Railroad Retirement Board v Fritz} can be analysed in a similar fashion and also demonstrates how this is an issue of rhetoric as well as substance.\textsuperscript{114} The law at issue granted dual railroad and social security retirement benefits for railroad workers with 10 years’ experience and some connection with the railroad in 1974, while not granting the same double

\textsuperscript{109} Union of Refugee Women (n 27 above) paras 37 - 42. Price argues that the purpose was in fact more specific: ‘The overall purpose of the legislation was to regulate the private security industry in a way that protects public safety. The specific purpose of the condition that people first be registered was to ensure that only trustworthy people work as security guards, for untrustworthy guards pose a risk to safety’ in Woolman & Bilchitz (n 7 above) 70. This train of thought is another good example of how the purpose of the same law can be stated at various levels. It also shows the overlap between purpose and means. Is ensuring that people are trustworthy the purpose of the section or is it the means to ensure public safety? Price is probably correct that the Union of Refugee Women majority did rely on one of the more specific purposes. But I do not think that undermines the point I make about how the failure to consider the additional requirements for registration made being a citizen or refugee redundant; even on the more specific purpose, there is no relation between means and end.

\textsuperscript{110} Union of Refugee Women (n 27 above) para 38.

\textsuperscript{111} I argued earlier that where a purpose is phrased at this level of generality it will be impossible for the law to fail rationality review. This example might seem to undermine that contention because I argue that the law could properly be found not to be rationally related to the general purpose of public safety. I do not think this conflicts with my earlier contention. The problem in this case is that government (or the Court) identified the wrong general interest. If they had identified the interest as ‘economic welfare’, the statute, even considering the additional 10 year requirement, would have been rational.

\textsuperscript{112} Union of Refugee Women (n 27 above) paras 118 - 119 (‘the purpose … seems to be amply covered by the other provisions of section 23 which achieve this purpose more appropriately and without discriminating against refugees’).

\textsuperscript{113} Another way of looking at this is that the 10-year requirement ‘flushes out’ the real purpose of the statute. The provision was based either on naked suspicion of refugees or was intended to economically advantage citizens and permanent residents.

\textsuperscript{114} n 17 above.
benefits to people with identical service but who had no railroad connections in 1974. Justice Rehnquist upheld the law, characterising the differentiation as between 'persons who had actually acquired statutory entitlement to windfall benefits while still employed in the railroad industry [and those] ... who were no longer in railroad employment when they became eligible for dual benefits'. Justice Brennan in dissent preferred this description: 'A retiree is favoured by retention of his full vested earned benefits if he had worked so much as one day for a railroad in 1974 ... [T]he fortuity of one day of employment in a particular year ... govern[s] entitlement to benefits earned over a lifetime.'

Judged in its overall context, the majority phrases the law as distinguishing on the basis of when a person became eligible for social security benefits. The dissent sees the classification as based on time. While this is partially a question of rhetoric – both approaches seem like accurate descriptions – the different way the classification is phrased seems largely determinative of the outcome.

The law can also be treated at different levels of abstraction. This technique is best demonstrated by cases where the government is defending a regulatory scheme that has the effect – intended or otherwise – of excluding a particular group. A comparison of two cases concerning the regulation of hairstylists and florists is instructive. In *Cornwell v Hamilton*, a district court held that California's requirement that African hair-braiders pass its cosmetology exam was not rationally related to a legitimate government purpose. It conducted a thorough examination of the activities in which the plaintiff hair-braiders engaged and the subject matter of the exam questions, and concluded that the plaintiff's activities made up only 11 per cent of the exam; not enough in its view to establish a rational relationship. Of course, the court perhaps required a stricter than usual degree of relationship between means and ends, but if it had not considered the detail of the exam and seen the law merely as requiring some sort of exam it might well have reached a different conclusion.

The failure to consider the specific questions in an exam is part of the explanation for the result in *Meadows v Odom*. The statute required all florists to take an exam before they could arrange flowers. Rather than

115 Fritz (n 18 above) 178.
116 Fritz (n 18 above) 196 - 197.
117 80 FSupp 2d 1101 (SD Cal 1999).
118 *Cornwell v Hamilton* (n 117 above) 1115.
119 *Cornwell v Hamilton* (n 117 above) 1108 1117 (‘Assume the range of every possible hair care act to involve tasks A through Z. From the Court's perspective, Cornwell's activities would cover tasks A, B, and some of C. The state's cosmetology programme mandates instruction in tasks B through Z. The overlap areas are B and part of C. This minimal overlap is not sufficient to force Cornwell to attend a cosmetology school in order to be exposed to D through Z, when she only needs B and a portion of C’).
120 360 FSupp 2d 811 (MD Louisiana) vacated as moot by 198 FedAppx 348 (5th Cir 2006).
striking down the law as facially absurd – as most reasonable people would have been tempted to do – the Court held that it furthered the important purpose of protecting public health because of testimony that ‘trained’ florists are ‘very diligent about not having an exposed pick, not having a broken wire, not hav[ing] a flower that has some type of infection, like, dirt that remained on it when it’s inserted into something they’re going to handle’. Presuming that was the legitimate purpose the law was meant to forward, one would have expected a showing that the exam actually included questions aimed at avoiding those injuries. However, because the court defined the law simply as having an exam without regard to the content of the exam, the law could be considered rational.

An example from South Africa shows the effect, in a slightly different context, of changing the level at which the law is defined. In 2005, the government re-arranged the borders of a number of provinces to abolish ‘cross-boundary municipalities’ – municipalities that had a provincial boundary running through them – that had proved extremely wasteful and inefficient. Two municipalities challenged the law in part on the basis that it was irrational for government to have placed them in province A rather than province B. Two Constitutional Court justices took very different approaches to this issue. Sachs J, although not deciding the issue, made it clear that, although the benefits of abolishing cross-boundary municipalities presented a legitimate goal, government also had to justify its decision to locate a municipality in one province rather than another.

In a later case, Van der Westhuizen J rejected that reasoning and defined the act requiring justification as the abolition of all cross-boundary municipalities. He viewed further interrogation of where the provincial boundaries eventually fell as a form of ‘second-guessing’ the legislature. The point of this example is as follows: If the law is viewed through Justice Sachs’s lens, then a mere resort to the purposes served by abolishing all cross-boundary municipalities – saving costs, better service delivery and so on – are insufficient because, in the absence of additional evidence, they have been equally well served by moving the municipality to the other province. A different or additional purpose is needed.

121 Meadows v Odom (n 20 above) 824.
122 The cases discussed here were not an equal protection challenge but were based on the general principle of the rule of law. However, the only reason for this was that the law which differentiated was a constitutional amendment and therefore not subject to conformity with the Bill of Rights. It was, however, subject to the more basic rationality norm implicit in the rule of law. If the change had been effected by normal law, an ordinary equality challenge could have been brought. I discuss it because I think it is the best example of the phenomenon.
123 Matatiele Municipality v President of the Republic of South Africa (1) 2006 5 BCLR 622 (CC) paras 101 - 108 (Sachs J concurring).
124 Merafong Demarcation Forum v President of the Republic of South Africa 2008 5 SA 171 (CC) para 114.
125 As above.
These examples indicate that, as with the previous three elements of the test, there is necessarily substantial room for judicial discretion in applying the rationality test. The way that a judge chooses to characterise the law can impact and even determine the outcome of the means-end test. Moreover, if the laws are defined too precisely – every exam question, or every decision as part of a broad scheme – government will become unmanageable. It seems a natural part of regulation of a complicated state that some actions which have a minimal impact might not have an explanation. If even the most minimal disadvantage or differentiation were to require scrutiny under the rational basis test, it would impinge on the important concerns about the separation of powers that underlie the rational basis test.

5 Rationality is dead

In the previous section I demonstrated that at each of the four points of the rational basis test there is room for judicial manipulation that can drastically alter the outcomes of the test. In this section I ask whether that malleability prevents the test from performing the task it is meant to serve.

The rational basis test is supposed to ensure that no law is passed that is motivated by a naked preference for one group of people over another, and that all laws that have the effect of disadvantaging a group of people also in fact serve, in some way, a public purpose. In order to achieve only those narrow goals the test uses the language and tools of rationality – rather than, for example, reasonableness – in order to prevent judges from second-guessing substantive legislative judgments about what policies can be pursued. The test

126 The direction of Price’s response to this paper is that rationality review is not in need of saving as the existing justifications are sufficient and it has been properly applied by the South African courts. I will admit that the structure of this article – breaking down rational review just to rebuild it – is somewhat rhetorical and that I tend to exaggerate the disease in order to shine the spotlight on the cure. However, while Price makes a compelling case, I remain unconvinced that rationality review – even in South Africa – is in perfect health. Nonetheless, I am happy to lose the descriptive battle about what the currently accepted justifications for rational basis review are, and how courts in fact apply the test today, if I win the conceptual/normative war about what the justifications should be and how the courts should apply it. Price and I are largely in agreement on the bigger issue, which I discuss in part 6.

127 Price argues that there is a clear distinction between rationality and reasonableness: Rationality requires that there is at least one reason for an action; reasonableness demands that the reasons for the decision outweigh the reasons against it. While he says we can imagine a spectrum that flows from rationality through to reasonableness, it is better, for the sake of certainty, to keep the two forever apart in Woolman & Bilchitz (n 7 above) 53 - 54. I agree entirely with Price on this. The failure to maintain clear boundaries between the standards of review is one of the major failings of American equal protection law, and it has lead to legal uncertainty and enlarged the space for judicial manipulation. However, to the extent that Price might be read to imply that my argument endorses such a spectrum (I do not think this was his intention, but some readers might interpret it that way), I wish to point out emphatically that it does not. I contend that, within the limits of the rational basis test, however defined – even in Price’s ‘one-reason-for’ formula – there is ample room for manoeuvre. Whether a single reason is a reason for a purpose depends on all the moveable parts I identify in part 4.
must meet all these purposes, it must outlaw the bias and arbitrariness while denying judges the ability to substitute their views for those of the legislature. My argument is that the test's acute malleability prevents it from reliably keeping these three balls in the air. In general, the test vacillates between being a rubber stamp for any government plan no matter how ludicrous, and a barrier to reasonable regulation with which judges disagree for other substantive reasons. Here I briefly recap the main problems with each part of the test.

Firstly, the ability to redefine naked preferences into legitimate purposes, combined with the inherent impossibility of objectively determining actual purpose, prevents courts from invalidating laws motivated by animus to a disfavoured group. Any law motivated by prejudice can be phrased as a law intended to – and which does in fact – serve some public purpose. As courts are rightfully unwilling, and practically and theoretically unable, to determine whether the real purpose is noble or evil, the test cannot reliably outlaw any law based solely on prejudice.128

A decision that a law is based on such malice rests on judicial judgment drawn from multiple factors. That renders the test an unreliable means to root out legislative prejudice. It is the kind of task that will perhaps inevitably be performed imperfectly, but if the rational basis test does not aid courts to identify and proscribe naked preference any further than to identify it as a goal, perhaps it is necessary to rethink either the test or what we expect it to achieve.

Secondly, the ability of government, advocates and the courts to manipulate the level of abstraction at which both the means and the end are defined converts the test into an exercise of judicial discretion. A court that wants to invalidate a law can do so by, for example, looking at the precise terms and effects of a law and phrasing the purpose relatively broadly so that it appears far removed from the means. While it is possible to outlaw purposes that are either tautologically specific or absurdly general, the grey area between those two extremes supplies sufficient space for courts to determine the outcome by the way they phrase the purpose. When it comes to identifying laws, there is no way to limit the courts to positing the law at a level that is neither too broad nor too specific. This has a tendency to undermine the requirement that courts do not second-guess legislative judgment. They can privilege their own discretion over that of legislatures while disguising that choice in the language of rationality.

Thirdly, the degree of connection required between means and ends. If the level of connection is set, as it usually is, as requiring the law to further the

---

128 Using the means-end part of the inquiry to expose prejudice does not remedy this deficiency. When a naked preference is rephrased as a public purpose, the law will achieve both equally well. It is only through suspicion that the legislature was really motivated by the impermissible purpose that the law can be struck down. But that requires an analysis of the law that is uncertain, intrudes substantially into the functioning of the legislature and is potentially fruitless.
purpose only in some conceivable way then all laws that identify a legitimate purpose – which we have seen is always possible – will pass the test. This standard fails to fulfill the requirement that a law must be connected to a public end because the connection is determined by the ingenuity of government and lawyers and the discretion of courts, not the effects of the law; any law can be saved by dreaming up some purpose that it might theoretically serve. It is only by raising this standard that rationality review can become meaningful. But once the standard is raised to something like ‘a fair and substantial relation’, the test affords judges too much power as many – maybe even a majority of laws – might reasonably be said not to ‘substantially’ advance their purpose. Courts have not been able to identify a mid-point that is both meaningful and restrained.

Fourthly, an element of the means-end test is the use of empirical evidence. If courts refuse to engage it at all, then the test – no matter what theoretical standard is employed – becomes an armchair game for lawyers and judges. While the possibility that a law may further a purpose might be deduced from ‘rational speculation’, the extent to which the purpose is advanced must rest on empirical facts. Evaluative rationality – the essence of rational basis review – only makes sense if courts are willing to examine empirical facts. But if the courts begin to evaluate the relative worth of competing empirical claims or to make judgments where empirical knowledge is uncertain, then we must ask: Why do we trust their judgment in that task more than that of the legislators?

The discretion inherent in the application of the test creates two types of problems. First, there is a massive gap between the rhetoric of courts which emphasises the mechanical nature of the test, and the degree of discretion courts in fact exercise. The lack of transparency is an evil in itself: Courts should practise what they preach or, more precisely, preach what they practise. It also reduces certainty. As the American experience demonstrates, courts can apply variable levels of scrutiny while claiming to adhere to the same standard. Moreover, it makes life difficult for litigants who cannot explicitly make some of the arguments that would influence courts, because courts officially deem those arguments irrelevant. Forcing those arguments into the subtext of litigation serves nobody. This is not a criticism that is unique to rationality review; it can and has been made about many areas of law, but it is especially blatant in the rationality context. Justice Marshall has explicitly commented on the discrepancy between word and deed, between rhetoric and reality:

129 Royster Guano (n 23 above) 415.
130 Beach Communications (n 20 above) 315.
131 See Bice (n 4 above) 13.
132 As above. (‘The formal elements of evaluation concern the logical pattern of the actor’s beliefs, namely whether the actor believes that his behaviour will achieve his goals and whether he believes that his actions will do so efficiently. The empirical elements of evaluation cannot be satisfied formally; they must be satisfied by reference to an objective standard outside the actor.’)
there are problems with deciding cases based on factors not encompassed by the applicable standards. First, the approach is rudderless, affording no notice to interested parties of the standards governing particular cases and giving no firm guidance to judges who, as a consequence, must assess the constitutionality of legislation before them on an ad hoc basis. Second, and not unrelatedly, the approach is unpredictable and requires holding this Court to standards it has never publicly adopted. Thus the approach presents the danger that … relevant factors will be misapplied or ignored.\textsuperscript{133}

Second, rational basis review does not achieve the ends it is meant to achieve. It is either so weak that it fails to invalidate the laws it is meant to invalidate or so uncertain that it treads on the ‘negative’ requirement underlying its existence: Courts must not interfere with legislative wisdom. Of course, many legal tests are imperfect. They often rely on the proper exercise of judicial judgment to achieve their supposed purposes: Very few (if any) legal tests are self-executing mechanisms. My criticism is limited to two points. One, understanding the detailed workings of the test suggests that rationality review serves an additional purpose – that of forcing government to justify its decision to treat people differently. Two, the test can be structured better to more closely achieve its goals – including the additional purpose of justification – than the current formulation achieves.

The next section explores the new rationale and shows how it can bridge the gap between the test and its goals. It then proposes a reformulation of the test that aims to acknowledge the malleability of the test and guide judges to best achieve the purposes of the test.

6 \textbf{Long live rationality!}

In \textit{Prinsloo v Van der Linde}, Justices Ackermann, O’Regan and Sachs, writing together, identified three pillars supporting rational basis review. I have mentioned two of them earlier, but to ensure I leave nothing out this time, I repeat the whole passage here:\textsuperscript{134}

In regard to mere differentiation the constitutional state is expected to act in a rational manner. It should not regulate in an arbitrary manner or manifest ‘naked preferences’\textsuperscript{135} that serve no legitimate governmental purpose, for that would be inconsistent with the rule of law and the fundamental premises of the constitutional state. The purpose of this aspect of equality is, therefore, to ensure that the state is bound to function in a rational manner. This has been said to promote the need for governmental action to relate to a defensible
vision of the public good,\textsuperscript{136} as well as to enhance the coherence and integrity of legislation.\textsuperscript{137}

\textit{In Mureinik’s celebrated formulation, the new constitutional order constitutes ‘a bridge away from a culture of authority . . . to a culture of justification’.}\textsuperscript{138}

The third, and crucial, purpose is the idea that rational basis review promotes a ‘culture of justification’. But what does that mean exactly? The phrase, as the Court notes, was coined by Mureinik, who defined it as:

a culture in which every exercise of power is expected to be justified; in which the leadership given by government rests on the cogency of the case offered in defence of its decisions, not the fear inspired by the force at its command. The new order must be a community built on persuasion, not coercion.\textsuperscript{139}

This particular articulation and emphasis is born from South Africa’s transition from past authoritarian government to a new constitutional order and has spawned an extensive South African literature on the notion of ‘transformative constitutionalism’.\textsuperscript{140} However, the idea of a culture of justification rests on the much more basic idea extremely familiar to American ears that courts act as forums of principle. As Dworkin puts it:

We have an institution that calls some issues from the battleground of power politics to the forum of principle. It holds out the promise that the deepest, most fundamental conflicts between individual and society will once, someplace, finally, become questions of justice. I do not call that religion or prophesy. I call it law.\textsuperscript{141}

The idea that the courts are a place where government must come to justify its treatment of its people is as much part of the American tradition as it is a tenet of the South African transformation.

Justification also links with the point I made earlier that rationality review must take place within a ‘public value’ model of politics.

\textsuperscript{136} Tribe (n 33 above) 1451.
\textsuperscript{137} P van Dijk & G van Hoof \textit{Theory and practice of the European Convention on Human Rights} (1990) 539.
\textsuperscript{138} E Mureinik ‘A bridge to where? Introducing the interim Bill of Rights’ (1994) 10 \textit{SAJHR} 31 32 cited in Makwanyane (n 32 above) para 156 n 1.
\textsuperscript{139} Mureinik (n 138 above) 32.
Justification emphasises not only that every law must serve a conceivable public good, but that the government must be able to articulate and defend its own vision of the good and how each law fits into that vision.

But what exactly does seeing rational basis as a requirement for government to justify its actions add to the existing rationales and how does it avoid the pitfalls identified earlier? It is not simply a requirement that the purpose of a statute be articulated by government’s lawyers in litigation. That would not solve the problem because the glitches in the other parts of the test would still render it hollow.\textsuperscript{142}

The appeal to justification demands that rationality review is not about trying to apply a mechanical test that resolutely denies responsibility and forces judges to make substantive choices. Justification makes the inquiry about \textit{persuasion} instead of logic.\textsuperscript{143} Persuasion depends on the observer to which it is addressed; the same case may persuade you but not me. That is not true of rationality; an action is rational or irrational independent of who asks the question. When we see rational basis review as representing part of the culture of justification or as part of the recognition of courts as forums of principle, it becomes a right for those who are disadvantaged by laws to demand that government explain the law and convince judges that the law serves some public purpose.\textsuperscript{144}

The benefit of the justification or persuasion rationale is twofold. It acknowledges that the extent of information and argument required in any case will differ vastly between judges – exactly what the history of the rational basis test suggests – and that the outcomes of cases are necessarily unpredictable.\textsuperscript{145} At the same time, it justifies the existence of this

\textsuperscript{142} Fritz (n 18 above) 197 - 198 (‘Equal protection rationality analysis does not empower the courts to second-guess the wisdom of legislative classifications. On this we are agreed, and have been for over 40 years. On the other hand, we are not powerless to probe beneath claims by Government attorneys concerning the means and ends of Congress. Otherwise, we would defer not to the considered judgment of Congress, but to the arguments of litigators.’)

\textsuperscript{143} Price finds the justification for rationality review in what he calls the promotion of ‘public reason’ in Woolman & Bilchitz (n 7 above) 60 - 62. I do not think that Price has something substantially different in mind when he uses ‘public reason’ from what I mean by ‘justification’ and ‘persuasion’. There may be some subtle differences, but the essence seems the same. In that case, we are largely in agreement about what the fundamental justification for rationality review should be. To my mind, most of our other debates should be seen as peripheral to this central point of agreement.

\textsuperscript{144} I should make clear that it is the judges who must be persuaded by the government. I do not think rationality requires the least well off to be persuaded that a law is rational. That would be a much stricter test as it would be unlikely to permit rational trade-offs. I thank James Fowkes for pointing this out to me.

\textsuperscript{145} Price argues that ‘the flexible, discretionary character of rationality review’ is not a negative, but ‘part of its value’ because it allows courts to apply the same test across a range of different contexts in Woolman & Bilchitz (n 7 above) 61. I agree. But as Price also notes, the flexibility is both a value and a danger. It does allow courts to tailor their approach to the peculiar circumstances of the case. But if we do not openly acknowledge the flexibility it allows them to cloak their actions in the objective language of rationality. The flexibility also increases uncertainty. I think that our project should be to consider how to structure the test so as to get the most benefits with the least danger. And I think Price would agree.
uncertainty by grounding the test not only in a substantive commitment to abolish certain laws combined with a healthy respect for the separation of powers, but also in a semi-procedural understanding of the role of courts in a constitutional system. For those who have no claim under any other right, rational basis is a means – a procedure – to air their grievances. The test has fuzzy borders and general guidelines, but no bright lines or clear answers.

It also solves the transparency criticism by forcing the substantive and empirical debates and the importance of the way the question is framed – which has for so long gone on beneath the surface – out into the open. Judges do not need to worry that they are exceeding their powers under the test if they weigh empirical evidence, or note that the level at which the test is phrased determines the outcome. Indeed, they are failing in their duties if they do not engage in these questions. In the past, the rhetoric of rationality both permitted and required that judges suppress the value judgments and empirical evaluations that are inseparable from a meaningful inquiry into a law’s rationality. Seeing rationality as focused on justification, on the contrary, permits and requires judges to explain all the implicit judgments they were previously required to hide. In the long run, this requires a shift not only in the justifications and structure of the test, but also the discourse that surrounds the test. The mere word ‘rationality’ inevitably tends to obscure the inner workings of the test. In the future, we might do better to abandon the word altogether.

Justification brings the test closer to its already existing purposes: outlawing prejudiced and irrational laws by stressing how uncertain those terms are. It makes the inquiry meaningful by giving it a basis from which to work: the ability of government to explain its actions.

I should make absolutely clear that I do not believe that changing the rationale of the test or adopting the new structure I propose will remove the discretion of judges or de-politicise the rationality inquiry. My point is that those are inevitable parts of the rationality inquiry that are in conflict with the supposed purposes and surrounding rhetoric of rationality. We cannot defuse that conflict and save rationality by trying to make it a mechanical test: that is an impossible goal. We can save it by re-imagining rationality as a tool to force government to justify its laws. My approach does not deny the substantive and discretionary elements of rationality review, it embraces them. I do, however, believe that this shift, and the detailed structure I propose, will, over time, make the application of the test more predictable as the implicit judgments become explicit.

The immediately apparent danger of shifting the onus to government is that it will place too great a burden on government and grant courts undue license to interfere with the business of government. This is a well-founded fear and is the reason why the discretion has been suppressed and denied for so long. However, I think it is possible to structure the test in a way that places an appropriate limit on courts.
6.1 What will the modified test look like?\textsuperscript{146}

The basic form of the test can remain, but it will require a number of modifications on the periphery. The important point is that the test must not be structured as a mechanical enterprise; clear thought and experience show that task is futile. It must be structured in a way that curbs undue judicial intrusion, acknowledges the malleability of the test and permits judges to strike down all laws that government is unable to convince them are motivated by and serve a valid purpose. I offer here one possible construction of the test that tries to meet these goals.

(1) Differentiation
The onus is on the litigant to show that the law differentiates between groups.

(2) Purpose
(a) Government must assert what the purpose(s) of the law are.
(b) The court must accept that assertion as the purpose(s) of the law, even if it believes the law serves other more important ends, unless:
   (i) The complainant can provide compelling evidence that the law was in fact motivated by naked preference. If so, the law is invalid. Or,
   (ii) The purpose is stated either so specifically or so abstractly that it automatically justifies the law. If so, the court should recast the purpose at a level so that, if presented with the right evidence, it could conclude that the purpose was not met.

(3) Law
The meaning and effect of the law must be understood in light of:
(a) the evil complained of by the plaintiff;
(b) the entire statutory scheme and any other relevant laws; and
(c) the need not to interfere too extensively with the details of government administration.

4 Connection
(a) The onus is on government to show a connection between the law and its purpose.
(b) The law must have more than a trivial or hypothetical connection to its purpose but it need not substantially or materially advance it.
(c) Empirical evidence about the effect of the law may be presented by either side and the evidence must be considered by the court. However, there is no duty on government or the complainant to do so.
(d) If there is no empirical evidence, the court should base its decision on:
   (i) the facts relied on by the government, if they are plausible; or

\textsuperscript{146} Price is not particularly enthusiastic about my expansive suggestion. He believes that it largely reflects what the Constitutional Court already does (n 7 above) n 169. It may be that some of the principles I list here are implicit in the Court’s approach (although many of them are definitely not), but few of them have been stated explicitly. The thrust of my argument is that we need to explicitly state the details of how we conduct rationality analysis. If the Court wants to take credit for implicitly developing the principles that is fine by me, as long as they clearly announce and adhere to them. I also note that, with one exception (see Price (n 7 above) n 43 (disputing the need for a clear burden on the state)), Price does not disapprove of the substance of the test I propose.
Rationality is dead! Long live rationality! Saving rational basis review

(ii) if those facts are not known, the plausible facts most favourable to government.

Other constructions might better represent the interests I have identified as fundamental to the inquiry, but let me briefly explain the reasons for the various elements of my construction.

The onus is on the plaintiff to identify the differentiation to ensure that rationality review does not become a vehicle for forcing the government to justify every piece of legislation. It is also part of the basic requirement that a litigant clearly state her case. This is a threshold requirement that should be satisfied in almost all cases. Its main purpose is simply to frame the rest of the inquiry.

One of the most important elements of the test is that the government must identify the purpose that the law is intended to serve. This fits directly into the justification rationale: the government must explain why it chose to treat different groups of people differently. It is precisely because the purpose of rationality review is to elicit a justification from government that the court must ordinarily accept the government’s asserted purpose, even if it believes that is not the true purpose. By forcing government to pin its colours to the mast, it also avoids the deficiency of the American jurisprudence that permits judges to imagine hypothetical purposes. It also avoids the immense evidentiary difficulties of judicial inquiries into legislative motive.

However, that is not a foolproof solution as the government is unlikely to admit to a prejudiced purpose and is likely to state the purpose at a level of abstraction that decides the rationality inquiry in its favour. Hence, there are two provisos to the general rule that government’s stated purpose must be accepted. Firstly, where there is clear evidence that the government was in fact motivated by naked prejudice, the complainant can produce that evidence which is determinative of the inquiry. Secondly, the court can recast the purpose at a different level of abstraction.

147 Price seems to suggest that courts should never consider legislative motive and points to the recent decision in Poverty Alleviation Network (n 12 above) as supporting this position (n 7 above) n 24. Price and I agree that it is generally fruitless to make a direct inquiry into legislative motive. He is also correct in saying that the Constitutional Court jurisprudence supports the position that motive is irrelevant. However, I do not believe that motive can always be ignored. Where there is clear evidence that the purpose which the government brings before the court is not the real purpose of the legislation but a construct designed to secure the law’s validity, courts should consider that information. These cases will be extremely rare: Legislators generally are motivated by admirable goals and, even when they are not, they are careful to hide their nefarious intentions. However, the importance of the proviso is both to guard against the rare case and to remind legislators to find legitimate reasons for all their decisions. The other important point is that rationality review (in both its equality and general guises) applies not only to legislative decisions made by multi-member bodies, but also to decisions taken by individuals. We should not allow the difficulties in determining motivation for large bodies to prevent us from inquiring into the motives of individuals.
to make the inquiry meaningful. In doing so, it should, as far as possible, retain the substantive content of the purpose.

The next step is defining the law. While the complainant must show some differentiation at the threshold, the exact nature of that differentiation is left to a later stage. It makes no sense here to place an onus on either party as both will, if possible, simply define the law in a way that suits their ends. The definition of the law will not always be open to manipulation, but where it is there seems to be no alternative other than to afford a court a guided discretion to decide what impact it is requiring the government to justify. In doing so, it should keep in mind both that rationality review must be meaningful and that government cannot be expected to defend every minute detail of every administrative scheme.

The core of the inquiry – the connection between the law and its purpose – is also the area that most evades a concise standard. It is perhaps this difficulty that has led judges to embrace the certainty of a standard that permits any conceivable connection to suffice. Yet experience – especially that of American courts – shows that such a standard either renders rationality review toothless, or masks the application of higher levels of scrutiny. The best solution seems to be to exclude the two extremes – any conceivable connection and substantial advancement – and require judges to seek for some degree of connection between those two poles, without trying to give it a label that is more likely to deceive than enlighten.

The other important part of the connection inquiry is the use of empirical evidence. The American courts’ shirking of empiricism in this context – they are perfectly happy to use it in other constitutional contexts148 – is another one of the causes of the test’s ‘dental’ deficiencies. It undermines the whole idea of requiring the state to act rationally for courts to permit it to act contrary to clearly established facts. However, placing a burden on either party to produce evidence would again be futile as in many cases there may not be reliable evidence available. That should not be the end of the case, but in the absence of facts, courts should give the government extensive leeway in determining what the true factual position is. If the government has clearly stated what it believes the factual situation is – and that assessment is plausible – the court should judge the law on that basis. Otherwise it should assume the plausible situation most favourable to government.

148 Most famously, see Brown v Board of Education of Topeka 347 US 483 n 11 (1954) (relying on sociological evidence about the impact of segregation on black children to support its decision to strike down the ‘separate but equal’ doctrine).
7 Conclusion

Despite the comparative detail of this test – it clearly lacks the easy-to-recite brevity of the original – it leaves many important questions unanswered. What is a ‘differentiation’? Even a seemingly universal law such as ‘theft is a criminal offence’ can be viewed as a differentiation between those who steal and those who do not, between rich and poor, between kleptomaniacs and those of us lucky enough not to have those urges. It also does not tell us exactly what sort of evidence would be needed to establish a nefarious subjective motive. Are statements in the legislative record enough? If so, what percentage of statements and by whom? Or must the illegitimate purpose appear in the statute itself? The test fails to answer at what time the connection between law and purpose must be made. Must it have been rational at the time the law was enacted, or the time the challenge was brought? Advances in empirical research or changes in social circumstances may alter the answer between those two points in time.

These are just some of the questions that I leave unanswered, partly for reasons of space, and partly because some of these problems do not admit of clear statement and are better solved through application in particular cases. I see the role of my proposed construction not as answering all the difficulties that inevitably arise in rationality inquiries. Its purpose is to structure the inquiry in a way that saves the constitutional requirement of rational state action from irrelevance, while at the same time guiding courts away from substituting their judgment for that of the legislature. Readers can decide for themselves whether my means are ‘rationally connected’ to my end ...
CHAPTER 2

THE CONTENT AND JUSTIFICATION OF RATIONALITY REVIEW

Alistair Price*

1 Introduction

It is a general principle of South African constitutional law that every law and every exercise of public power should not be arbitrary but, instead, should be rational.¹ This principle, developed by the Constitutional Court in a series of judgments,² sets rationality or non-arbitrariness as a necessary condition that every law or act of all branches and spheres of the state must satisfy in order to be legally valid. It operates as a minimum standard – a constitutional baseline – that applies even in circumstances where no fundamental right or other constitutional standard is directly applicable. By developing this principle, the courts have asserted a significant power, for they are now able to assess, on substantive grounds, the rationality of every provision in the statute book and all conduct of the executive, public administration, and other organs of state. They do so, in general, by deciding whether the law or conduct in

¹ Throughout this essay, I refer to this general prohibition on arbitrariness as the ‘rationality principle’ and describe its application by the courts as ‘rationality review’.

² S v Makwanyane 1995 3 SA 391 (CC) para 156; Prinsloo v Van der Linde 1997 3 SA 1012 (CC) para 24; New National Party v Government of South Africa 1999 3 SA 191 (CC) paras 19 & 24; Pharmaceutical Manufacturers Association of South Africa: In re Ex parte President of South Africa 2000 2 SA 674 (CC) paras 85 & 90; United Democratic Movement v President of South Africa (No 2) 2003 1 SA 495 (CC) para 55; Affordable Medicines Trust v Minister of Health 2006 3 SA 247 (CC) paras 74 - 79; Albutt v Centre for the Study of Violence and Reconciliation [2010] ZACC 4 para 49; Poverty Alleviation Network v President of South Africa [2010] ZACC 5 para 65; Law Society of South Africa v Minister for Transport [2010] ZACC 25 paras 32 - 39; and Glenister v President of South Africa [2011] ZACC 6 paras 55 - 70. For a comparative analysis of Poverty Alleviation Network and Albutt, see A Price ‘Rationality review of legislation and executive decisions’ (2010) 127 SALJ 580 - 591. Due to time and spatial constraints, I have been unable to include a discussion of Law Society of South Africa and Glenister. Fortunately, the reasoning and outcome of those decisions is consistent with, and in many respects confirms, the arguments advanced below.
question is rationally connected to a legitimate government purpose. If that minimum standard is not met, the law or conduct is unconstitutional. The courts have held that this power is justified by the rule of law, which, by virtue of section 1(c) of the Constitution, is a founding value of South African law. Without doubt, the uniquely broad applicability of this rationality principle, and its apparently foundational nature as an entailment of the rule of law, underline the importance of a sound grasp of its content and justification.

Bishop has written an insightful and provocative essay exploring rationality review only as it applies to legislative differentiations between classes of persons in terms of section 9(1) of the Constitution. For our courts apply that provision by invalidating legislative differentiations that are irrational or arbitrary.

Bishop’s argument has two legs: the first critical; the second constructive. He argues that the structure of rationality review, and perhaps also the manner in which it has been applied, ensure that it cannot serve its traditional justifications. He then proposes an additional justification for the rationality principle, as well as a new structure for its application, which together, he claims, ‘save’ rationality review. What follows here is an attempt, in a spirit of collaboration, to evaluate Bishop’s argument. I do so indirectly, by offering an alternative account of the content and justification of rationality review that is both broader and narrower than his. It is broader because I consider rationality review of legislation and exercises of public power generally, not only of differentiations in terms of section 9(1). It seems to me that this broad applicability, together with the courts’ view that rationality review is justified by the rule of law, call for a wider analysis. My account is narrower because it focuses solely on South African law, whereas Bishop compares rationality review in South Africa and the United States and draws conclusions that seem to apply equally to both countries.

I am hesitant to follow Bishop down the comparative road for two reasons. First, rational basis review under the US equal protection clause applies to many legislative differentiations that in South Africa are tested against the more rigorous prohibition on unfair discrimination in terms of

---

3 The content of the rationality principle is analysed more closely below, from 7.
4 See, eg, Prinsloo (n 2 above) para 25; City Council of Pretoria v Walker 1998 2 SA 363 (CC) para 137; New National Party (n 2 above) para 24; Pharmaceutical Manufacturers (n 2 above) para 85; Albutt (n 2 above) para 49; and Poverty Alleviation Network (n 2 above) para 65.
5 M Bishop ‘Rationality is dead! Long live rationality! Saving rational basis review’ in S Woolman & D Bilchitz (eds) Is this seat taken? Conversations at the Bar, the bench and the academy about the South African Constitution (2012) 1. Sec 9(1) provides that ‘[e]veryone is equal before the law and has the right to equal protection and benefit of the law’. Bishop suggests that his conclusions may apply beyond the context of sec 9(1).
6 This rule was first established in Prinsloo (n 2 above) para 24.
7 The relevant part of the Fourteenth Amendment to the US Constitution provides that ‘[n]o state shall … deny to any person within its jurisdiction the equal protection of the laws’.
The content and justification of rationality review

section 9(3) of the Constitution. For example, differentiations on grounds of sexual orientation, age, or HIV status need merely be rational in the United States, whereas South African courts assess whether they constitute unfair discrimination. Some of the academic disapproval of rational basis review in the United States may be partly motivated by the relative ease with which such differentiations are upheld as constitutional, because rational, in that country. In South Africa, by contrast, the broader reach and higher standard of the prohibition on unfair discrimination ensure that many of these differentiations will not survive constitutional scrutiny, whether or not they satisfy the rationality principle. Accordingly, rationality review arguably has costs in the United States that it does not incur in South Africa. Secondly, the justification of rationality review depends to a significant extent on the manner in which its evaluative and vague elements are applied by judges in particular cases. But it is by no means clear that since 1994 South African judges have applied the rationality principle in a manner identical to US judges. Whatever the record of the US courts may be, we need not accept Bishop’s claim that the South African Constitutional Court is guilty of applying the rationality principle in an inappropriate way. For these two reasons, we should be hesitant to tar rationality review in both legal systems with the same critical brush.

Having explained the difference between the scope of my account of rationality review and Bishop’s, I now outline the argument that follows. Part 2 surveys the different contexts in which the rationality principle applies. I then analyse its content, drawing attention to some of its characteristics – in particular, the fact that its application often requires courts to make discretionary judgments – which call for further explanation and justification. Part 3 takes the first step by analysing the distinction between reasonableness and rationality in constitutional law. It then explains the court’s duty to respect, or not to ‘second-guess’, the political autonomy of the executive and legislature, as well as the idea that judicial scrutiny can vary in intensity. This part concludes by briefly considering the notion of arbitrariness. Having explained the content of the rationality principle, I turn to its justification. Part 4 advances some reasons for believing that rationality review is justified in principle, relying prominently on the idea that its discretionary character is part of its value. Part 5 argues that rationality review is justified in practice, because the Constitutional Court has, on the whole, applied the rationality principle sensibly. My ultimate conclusion, then, is that rationality review does not need to be saved.

8 Sec 9(3) provides that ‘[t]he state may not unfairly discriminate directly or indirectly against anyone on one or more grounds, including race, gender, sex, pregnancy, marital status, ethnic or social origin, colour, sexual orientation, age, disability, religion, conscience, belief, culture, language and birth’.
9 See, for a classic account, L Tribe American constitutional law (1988) ch 16.
2 The content of rationality review

2.1 Contexts in which the constitutional principle of rationality is applicable

The first category of cases in which the rationality principle is applied are those addressing whether legislative differentiations between classes of persons comply with the right to equality before the law in terms of section 8(1) of the interim Constitution (from 1994 to 1996) and section 9(1) of the final Constitution (from 1997). In *Prinsloo v Van der Linde*, the majority held:

In regard to mere differentiation [ie differentiation that does not constitute ‘unfair discrimination’] the constitutional state is expected to act in a rational manner. It should not regulate in an arbitrary manner or manifest ‘naked preferences’ that serve no legitimate purpose, for that would be inconsistent with the rule of law and the fundamental premises of the constitutional state. The purpose of this aspect of equality is, therefore, to ensure that the state is bound to function in a rational manner ... Accordingly, before it can be said that mere differentiation infringes [the right to equality before the law] it must be established that *there is no rational relationship between the differentiation in question and the governmental purpose which is proffered to validate it*.

That case provides an instructive example. Section 84 of the Forest Act provided that negligence is presumed in delictual actions for damages arising from forest fires on land situated outside designated ‘fire control areas’. Prinsloo challenged the constitutionality of that provision when Van der Linde sued him for delictual damages allegedly suffered on his farm by the spread of fire from Prinsloo’s neighbouring land which was not located in a fire control area. The issue was whether the legislative differentiation between owners of land outside controlled areas, who were presumed negligent, and owners of land within such areas, who were not, was consistent with the right to equality before the law. The majority upheld the law on the ground that it was rationally related to the legitimate purpose of reducing the risk of fires spreading from land in non-controlled areas. Unlike landowners in controlled areas, the owners of non-controlled land bore no statutory duties relating to fire prevention; instead, the presumption of negligence provided an incentive to be vigilant. The presumption also alleviated the difficulty of proving fault, given that how a fire starts on neighbouring land, or what steps are taken to prevent or contain a fire, are generally within the peculiar knowledge of the neighbouring landowner. Accordingly, the differentiating

---

10 Although sec 9(1) (n 5 above), includes the additional concept of ‘equal benefit’ of the law, it was held in *National Coalition for Gay and Lesbian Equality v Minister of Justice* 1999 1 SA 6 (CC) para 59 that it should be applied exactly as sec 8(1) had been, that is, by way of rationality review.

11 *Prinsloo* (n 2 above) paras 25 - 26 (footnote omitted) (my emphasis).

law was not arbitrary and thus did not violate the right to equality before the law.

After Prinsloo, the rationality principle has consistently been applied to challenges in terms of section 9(1) to differentiating legislation or policy. In New National Party v Government of South Africa, however, the Constitutional Court extended its scope by holding, in effect, that all legislative schemes must bear a rational relationship to the achievement of a legitimate government purpose and that Parliament ‘cannot act capriciously or arbitrarily’. In that case, a political party challenged provisions in the Electoral Act which required possession of a bar-coded ID document to register as a voter and to vote. Writing for the majority, Yacoob J drew a distinction between rationality and reasonableness as standards of review of legislative schemes, holding that the former was the appropriate standard in the light of the separation of powers, and was justified because arbitrariness is contrary to the rule of law; the latter was relevant only to whether a violation of the right to vote under section 19 of the Constitution was justified in terms of section 36. He then held that the provisions were neither arbitrary nor capricious, because they were rationally related to the legitimate government purpose of ensuring the effective, efficient and reliable exercise of the right to vote.

The most important decisions are Prinsloo (n 2 above); Harksen v Lane NO 1998 1 SA 300 (CC); Walker (n 4 above); Jooste v Score Supermarket Trading (Pty) Ltd 1999 2 SA 1 (CC); Bel Porto School Governing Body v Premier of the Western Cape 2002 3 SA 265 (CC); Van der Merwe v Road Accident Fund 2006 4 SA 230 (CC); Union of Refugee Women v Director, Private Security Industry Regulatory Authority 2007 4 SA 395 (CC); Nyathi v Gauteng MEC of Health 2008 5 SA 94 (CC); Weare v Ndebele NO 2009 1 SA 600 (CC). S v Ntuli 1996 1 SA 1207 (CC), which was decided before Prinsloo, is not a genuine example of rationality review. There it was held that legislation requiring that unrepresented prisoners acquire a certificate from a High Court judge in order to appeal their Magistrate’s Court convictions in the High Court violated the right to a fair trial under sec 25(3)(h) of the interim Constitution and the right to equality before the law under sec 8(1). The thrust of the Court’s reasoning was that, unlike the leave to appeal procedure in the superior courts, the certification procedure was unsystematic, haphazard and over-broad, because it was not statutorily guided and was conducted by unrepresented persons. The consequent danger that meritorious appeals would not be heard violated sec 25(3)(h). Sec 8(1) was also violated because the violation of sec 25(3)(h) affected only one class of convicted persons, but not all. However, the Court certainly did not consider whether the violation of the fair trial rights of some prisoners but not of others was rationally connected to a legitimate government purpose!

n 2 above, para 19.
Act 73 of 1998.

n 2 above, para 24. In a dissenting judgment, O’Regan J held at para 122 that reasonableness, not rationality, was the appropriate standard to be applied in the context of the right to vote.

New National Party (n 2 above) paras 26 & 48. It is difficult to avoid the conclusion that, despite clearly distinguishing between the requirement of rationality (under the rule of law) and the right to vote, the majority nevertheless conflated these separate requirements when applying the law to facts. Yacoob J effectively held that the right to vote was not limited because the law was rational. But even if the requirement to possess a bar-coded ID was rational, why did it not nevertheless at least limit the right to vote of those unable to obtain a bar-coded ID? They were prevented from voting. Such a limitation, of course, would have required independent justification under sec 36 at the standard of reasonableness.
Since New National Party, the rationality principle has been applied to statutes outside the context of sections 9(1) and 19 on several occasions.\(^\text{18}\) In *United Democratic Movement v President of South Africa (No 2)*,\(^\text{19}\) the Constitutional Court unanimously rejected an argument that a legislative scheme permitting members of Parliament and the provincial legislatures to change parties without losing their seats during prescribed periods was arbitrary. The Court held that legislators’ motives when voting for legislation were irrelevant to rationality review,\(^\text{20}\) and that the scheme was rationally connected to the legitimate purpose of allowing floor-crossing – a purpose specifically contemplated by the Constitution itself.\(^\text{21}\) In *Merafong Demarcation Forum v President of South Africa*,\(^\text{22}\) the Court applied the rationality principle to those provisions of the Constitution Twelfth Amendment Act\(^\text{23}\) that purported to alter provincial boundaries so that the Merafong City Local Municipality, which initially lay partly in Gauteng and partly in the North West province, would be located entirely within North West. Remarkably, all the justices appeared to agree that it was appropriate to apply the rationality principle, not only to the provisions of the Act, but also to the prior decision of the Gauteng provincial legislature to approve the Bill, as part of the legislative process in terms of sections 74(3)(b) and (8) of the Constitution.\(^\text{24}\) Both

\(^{18}\) Further examples, in addition to those discussed in the text, are *Affordable Medicines* (n 2 above) paras 74 - 77, 96 - 100; *Poverty Alleviation Network* (n 2 above) paras 64 - 76; *Law Society of South Africa* (n 2 above) paras 32 - 33; *Glenister* (n 2 above) paras 55 - 70.

\(^{19}\) *n 2 above.

\(^{20}\) *United Democratic Movement* (n 2 above) para 56.

\(^{21}\) *United Democratic Movement* (n 2 above) paras 57-5 & 69 - 75. The enactment of floor-crossing legislation is envisaged by item 23A(3) of Annexure A to Schedule 6 of the Constitution.

\(^{22}\) 2008 5 SA 171 (CC). *Poverty Alleviation Network* (n 2 above) concerned a rationality challenge to very similar legislation and the Court followed its decision in *Merafong*.

\(^{23}\) Of 2005.

\(^{24}\) The dissenting justices, Mosevene DCJ, Madala and Sachs JJ, all agreed that the decision of the provincial legislature was irrational, because it acted ‘without a proper appreciation of its [Constitutional] powers and duties’ (para 124). Van der Westhuizen J doubted in his joint-majority judgment whether this extension of rationality review was appropriate (paras 72 - 74), but nevertheless concurred with the joint-majority judgment of Ngcobo J, who held that although it was generally undesirable for courts to speculate about what reasons move legislators to enact laws, the decision of the Gauteng legislature in question was different. It could appropriately be reviewed for rationality, because it was not ordinary legislating and the record contained a committee report setting out some reasons for supporting the Bill which the legislature ‘adopted’ by majority vote (paras 254 - 258).

This extension of rationality review is problematic, for two reasons. First, it is not merely difficult, but impossible, to tell what reasons subjectively moved legislators to vote for or against a Bill. As a collective body, there is no single set of reasons that subjectively moves all the legislators when they vote. That is so even where, as here, the majority of the legislature adopts a document setting out the reasons for a Bill which satisfied a legislative committee. For the committee and the legislature are different bodies and legislators are not constrained to follow the reasoning of committee members. This is borne out by Mosevene DCJ’s observation (para 176) that there were four possible sources of reasons for the Gauteng legislature’s decision to approve the Bill, which were not all consistent. What justified his relying only on one source of reasons, while disregarding the others (para 176)? The extension also appears to conflict with the unanimous ruling in *United Democratic Movement* (n 2 above) that legislative motives are irrelevant to rationality review (para 56). Secondly, it is impossible to conceive of some non-fictional, objective set of reasons that moved every legislator in his or her legislative voting. It surely follows that empowering courts to review
were upheld as rational.25

The rationality principle is also applied as a minimum standard to the conduct of the executive branch of the state. In the Pharmaceutical Manufacturers case,26 it was held:

It is a requirement of the rule of law that the exercise of public power by the executive and other functionaries should not be arbitrary. Decisions must be rationally related to the purpose for which the power was given, otherwise they are in effect arbitrary and inconsistent with this requirement. It follows that in order to pass constitutional scrutiny the exercise of public power by the executive and other functionaries must, at least, comply with this requirement. If it does not it falls short of the standards demanded by our Constitution for such action. The question whether a decision is rationally related to the purpose for which the power was given calls for an objective enquiry.27

This formulation makes it clear that no exercise of public power by the executive or other functionaries (presumably including 'organs of state' under section 239 of the Constitution and perhaps private bodies performing public functions) is beyond rationality review by the courts. The source of the court’s power in this regard is said to be the constitutional principle of legality, which is justified by the broader value of the rule of law, and has been held to require that public powers should not be misconstrued and should be exercised rationally and in good faith.28

Accordingly, in Pharmaceutical Manufacturers it was held that a presidential proclamation specifying the commencement date for national legislation dealing with human and animal medicines was irrational, because it provided for the statute to come into force before various schedules of substances and regulations essential for its operation had been made. Although the proclamation was made in good faith, it was arbitrary and, for that reason, was declared invalid.29 Similarly, in Kruger v President of South

the reasons that move legislators, especially in multi-stage legislative processes that include a series of collective decisions, embroils judges too deeply in the exercise of legislative power. The better approach is to apply the rationality principle only to the content of legislation, not to prior decisions to enact that legislation taken by legislators during the legislative process. Fortunately, the Constitutional Court seems recently to have adopted this view in Poverty Alleviation Network (n 2 above) paras 72 - 73. But that judgment does not disapprove or even discuss the different approach taken in Merafong. So it would seem that there are now two conflicting approaches to the same question in the law reports.

25 I discuss the challenge to the content of the Act below at 16 & 26.
26 n 2 above.
27 Pharmaceutical Manufacturers (n 2 above) paras 85 - 86.
28 FedSure Life Assurance Ltd v Greater Johannesburg Transitional Metropolitan Council 1999 1 SA 374 (CC) paras 56 - 58; President of South Africa v South African Rugby Football Union 2000 1 SA 1 (CC) para 148; Pharmaceutical Manufacturers (n 2 above) above para 85; Albutt (n 2 above) para 49.
29 Pharmaceutical Manufacturers (n 2 above) para 94.
Chapter 2

Africa, a presidential proclamation of amendments to the Road Accident Fund Act was invalidated on grounds of irrationality, because it mistakenly purported to amend a random selection of that Act’s provisions. Perhaps the most striking extension of rationality review into the executive domain is the suggestion in Kaunda v President of South Africa that the courts are empowered to review foreign policy decisions regarding requests for diplomatic protection and to set them aside if they are irrational. Finally, the Constitutional Court recently held in Albutt v President of South Africa that a procedural right to be heard can, on occasion, be based on the principle of rationality: The Court declared irrational the President’s refusal to hear victims of crimes committed by persons applying for a presidential pardon on the ground that they had acted with a political motive.

Rationality is also an important requirement of administrative law. In terms of section 33 of the Constitution, all administrative action must be reasonable, and rationality is a necessary (but usually insufficient) condition for reasonableness.

In addition, section 6(2)(f)(ii) of the Promotion of Administrative Justice Act (PAJA) provides that courts are empowered to set aside an administrative act if it is:

- not rationally connected to –
  - (aa) the purpose for which it was taken;
  - (bb) the purpose of the empowering decision;
  - (cc) the information before the administrator; or
  - (dd) the reasons given for it by the administrator.

Section 6(2)(e)(vi) provides for review of administrative action taken ‘arbitrarily or capriciously’ (although it is difficult to imagine an irrational administrative decision that would not also be arbitrary, and vice versa).

But PAJA applies only to ‘administrative action’ as defined in section 1. In contrast, the constitutional principle of rationality, based on the principle of legality and the right to equality before law, applies far more widely. Stretching beyond the courts’ traditional scrutiny of the

---

30 2009 1 SA 417 (CC).
31 Act 56 of 1996.
32 2005 4 SA 235 (CC) paras 77 - 80.
33 Albutt (n 2 above). When exercising public power, a procedural duty hear will be owed under rationality review where affording a hearing is necessary (ie ‘the only rational means’: paras 69, 90) to serve the legitimate purpose of the power. In this case, the Court held that hearings were necessary for the impugned pardons process to serve its professed purposes of promoting national unity and reconciliation. For further analysis of this decision, see Price (n 2 above).
34 C Hoexter Administrative law in South Africa (2007) 307, citing the use by O’Regan J of the formula ‘reasonableness or rationality’ in Bato Star Fishing (Pty) Ltd v Minister of Environmental Affairs 2004 4 SA 490 (CC) para 43.
administration’s acts, decisions and rules, this judicial power now extends deeply into the legislative process and executive domain. The decision as to whether a legislative or executive differentiation complies with section 9(1) of the Constitution is but one of the many areas in which it is applied. Moreover, in all these contexts, the courts frequently seek to justify rationality review by appealing to the rule of law.

2.2 The various forms of rationality review

There are subtle variations in the content of the constitutional principle of rationality. The ‘central case’, or exemplar, of rationality review is the test applied to legislation in general, which requires that the impugned law be rationally connected to a legitimate governmental purpose. The test applied to exercises of public power is slightly different: it requires that the impugned act or decision be rationally related to the purpose for which the power was given.

These tests are closely analogous to those under sections 6(2)(f)(ii)(aa), (bb) and (dd) of PAJA, which respectively require that the impugned administrative action bear a rational connection to the purpose ‘for which it was taken’, the purpose ‘of the empowering provision’ and the ‘reasons for it given by the administrator’. Less closely analogous, however, is the fourth rationality test under PAJA, namely the requirement that administrative action be rationally connected to ‘the information before the administrator’ (section 6(2)(f)(ii)(cc)).

Notwithstanding the conceptual overlap and subtle differences between these standards, the remainder of this essay focuses only on the constitutional principle of rationality as developed by the Constitutional Court, and specifically on its central case – that is, the requirement that legislation and executive conduct be rationally connected to legitimate government

35 See, eg, Merafong (n 22 above).
36 See, eg, Kaunda (n 32 above).
37 See, eg, the cases cited (n 4 above).
38 See, eg, Prinsloo (n 2 above) para 25; Harksen (n 13 above) para 53; Jooste (n 13 above) para 17; Van der Merwe (n 13 above) para 42; Weare (n 13 above) para 46.
39 Pharmaceutical Manufacturers (n 2 above) para 85. Albutt (n 2 above) illustrates that if a public power was not given for any particular purpose, then its exercise must be rationally connected to the purpose which the defendant intended to serve by invoking it (provided, of course, that that purpose is ‘legitimate’): in other words, the central case test is applied. In that case, the President sought to exercise his power to pardon criminals under sec 84(2)(j) of the Constitution specifically in order to promote national unity and reconciliation; accordingly, the manner in which he exercised it had to be rationally related to those purposes.
40 Presumably, if the ‘purpose for which it was taken’ or the ‘reasons given for it by the administrator’ are different to the ‘purpose of the empowering provision’, then the court will also have to decide whether the administrative action was based on an ulterior purpose or motive or ‘reason not authorised by the empowering provision’ in terms of sec 6(2)(e)(i) and (ii).
41 I do not assess rationality review cases under PAJA and sec 33 of the Constitution. A full assessment of the rationality review of ‘administrative action’ calls for a far wider study of High Court and Supreme Court of Appeal judgments which is not undertaken here.
purposes – for this version most clearly exemplifies the features of rationality review which stand in particular need of explanation and justification. In addition, since the other tests (the ‘peripheral cases’) share some but not all the features of the central case, arguably they can only be fully understood in the light of it.

2.3 The central case of rationality review

The precise meaning of the principle that law and exercises of public power must be rationally connected to a legitimate governmental purpose is far from clear. What follows is a constructive analysis or interpretation of this principle which seeks to advance our understanding of it. The primary source of confusion arises from the metaphorical requirement of a ‘rational connection’ or ‘relationship’ between the law and a purpose. What is meant by this? It seems to me that what rationality review calls for, at least in effect if not at the level of legal doctrine, is a two-pronged assessment of (1) a law’s purpose and (2) its actual or probable effect.

First, the purpose requirement: The court must decide whether the impugned law is intended to serve at least one legitimate purpose, for a law without a legitimate purpose is irrational. This in turn requires judges both to discern the purpose of a law and then to judge whether it is legitimate. Determining the purpose of the law in question is often straightforward, because it raises everyday questions of statutory interpretation and the respondent state can be expected to submit to the court what it considers the relevant purpose to be. It is less clear how judges are to decide whether that purpose, once discerned, is ‘legitimate’. The Constitutional Court has not explicitly addressed the issue, presumably because it has seldom proved contentious in practice. Clearly an evaluative judgement is called for. One approach is to hold that the court is free to approve or disapprove of the purposes endorsed by the executive and/or legislature by making an entirely open-ended value judgment. But it is doubtful whether courts should make such unconstrained judgments. A preferable alternative is to hold that a legal rule’s purpose is legitimate, under the rationality principle, if and only if it is consistent with all other legal and constitutional constraints, including the Bill of Rights and the Constitution’s ‘objective, normative value system’.

---

42 Both New National Party (n 2 above) para 24 and Merafong (n 22 above) para 63, eg, make it clear that there need only be a single legitimate purpose.
43 If the applicant can advance plausible reasons casting doubt on the rationality of the law or conduct, that is sufficient to oblige the state to explain why the law or conduct is in fact rational, and where necessary, to provide evidence supporting this assertion. Bishop (n 5 above) 33 - 34, argues that the courts should develop a legal rule placing a full onus on the state to establish rationality. This step is probably unnecessary in practice, given that the state will shoulder an explanatory and evidentiary burden as soon as the applicant advances a prima facie case. This threshold requirement is also arguably necessary to prevent vexatious and costly rationality review litigation.
44 A brief survey of the relevant case law, below at 24 - 25, suggests as much.
45 Carmichele v Minister of Safety and Security 2001 4 SA 938 (CC) paras 54 - 55.
This approach constrains judges to deciding only questions of law (Is this purpose unconstitutional? Is this purpose otherwise unlawful?), and clearly rules out certain purposes (eg unfairly discriminatory purposes and criminal purposes). Where government purposes are otherwise constitutional and lawful, courts should be loath to declare them ‘illegitimate’ under the rationality principle.

Secondly, the effect requirement: If the law indeed has a legitimate purpose, the court must then decide whether the law serves that purpose sufficiently well to merit the description of bearing a ‘rational connection’ or ‘relationship’ to it. It should immediately be apparent how vague the effect requirement is, because it specifies neither how, nor how well, the law in question should serve its purpose. Nevertheless, we may observe the following. A law intended to serve a legitimate end is undoubtedly arbitrary if it fails to serve that end at all. For example, the presidential proclamation in *Kruger*, which purported to amend and update the Road Accident Fund Act, was plainly intended to serve important ends, but failed to do so because it inadvertently referred to the incorrect sections of that Act. In contrast, a law will satisfy the effect requirement if it achieves its end symbolically (ie has intrinsic value), or if it is a means that tends to bring about its end as a matter of empirical causality (ie has instrumental value). An example of the former is a legal rule obliging radio stations to play a minimum number of hours of South African music every week, which is rational because it symbolically achieves the legitimate end of supporting and promoting South African culture. An example of the latter is provided by *Prinsloo*, where the Court effectively approved the state’s empirical assumption that the presumption that landowners outside fire-controlled areas are negligent in respect of fires spreading from their property was likely to alter their behaviour in a way that would reduce the risk of fires.

So, in sum, the rationality principle effectively requires courts to apply an evaluative purpose requirement and a vague effect requirement. This analysis brings into sharp relief the fact that it often demands discretionary judgment. As Bishop has explained in detail, there are at least three ways in which judges exercise discretion when applying the rationality principle. First, judges have freedom to choose whether to assess the challenged legal rule in isolation or as part of a broader legislative scheme. This is significant, for a rule considered in isolation may seem arbitrary or purposeless and yet constitute a vital cog in a larger legislative machine. For example, the presumption of negligence in *Prinsloo* was arguably intelligible only within the broader context of a scheme designed to reduce the risk of forest fires. Secondly, deciding whether a law has a legitimate purpose necessarily requires evaluative judgment, for the court must both specify the purpose at a meaningful

---

46 n 30 above.
47 Of course, a law will satisfy the effect requirement if it has both intrinsic and instrumental value.
48 n 2 above.
49 Bishop in Woolman & Bilchitz (n 5 above) 10 - 25.
level of abstraction and then determine whether, at that level of abstraction, it is legitimate. Finally, as mentioned, the vagueness of the requirement that the law serve its purpose sufficiently well to be ‘rationally connected’ to it often involves judicial choice, because it is unclear how, or how well, the law is required to serve its purpose.

This ‘malleability’ of rationality review – the fact that the application of the rationality principle often involves the exercise of judicial discretion – stands in tension with its second general feature, namely, the court’s frequent statements that rationality review does not involve an assessment of the reasonableness of legislation, nor its political merits or demerits, nor the making of policy choices, nor the substitution of the court’s opinions as to what is correct or appropriate for the opinions of the legislature or executive as the case may be. These statements reflect the courts’ general duty to respect the competence and political autonomy of other branches of the state. It is sometimes referred to as the requirement that the courts do not ‘second-guess’ legislative and executive choices.

These observations raise several important questions. First, how does an assessment of the rationality of law or conduct differ from an assessment of its reasonableness? Secondly, how can the discretionary character of rationality review be reconciled with the courts’ duty to respect the autonomy of the political branches of the state? And thirdly, is the courts’ frequent claim that rationality review is justified by the rule of law defensible? Any adequate explanation of the content and justification of rationality review must address these questions.

3 Reasonableness, rationality and arbitrariness

Before we can consider whether rationality review is justified, we need a firmer grasp of its nature. That depends in part on contrasting the constitutional standard of rationality with the related notions of reasonableness and arbitrariness. In this part, I attempt to shed some light on these distinctions in a way that explains the courts’ duty, introduced above, not to ‘second guess’ the choices of the political branches of the state when applying the rationality principle, as well as the different ways in

50 Eg, when applying the rationality principle to the presumption of negligence in Prinsloo (n 2 above), if the court had set that law’s purpose too abstractly (eg to serve the common good) or too specifically (eg to presume the negligence of neighbouring landowners in respect of forest fires), then rationality review would have become futile.
51 Bishop (n 5 above) 26.
52 New National Party (n 2 above) para 24.
53 Merafong (n 22 above) para 63.
54 Jooste (n 13 above) para 17.
55 Pharmaceutical Manufacturers (n 2 above) para 90; Bel Porto (n 13 above) para 45; Kaunda (n 32 above) para 79; and Merafong (n 22 above) para 63.
56 See, eg, Merafong (n 22 above) paras 114, 171 and Bishop (n 5 above) 9.
which judicial scrutiny of law and state conduct may be said, in the familiar metaphor, to ‘vary in intensity’.

3.1 Reasonableness and rationality

The appropriate point of departure is the notion of reasonableness. Simply put, an act is reasonable, in everyday life, if the reasons in its favour defeat the reasons militating against it. So to judge an act’s reasonableness involves an evaluation of competing considerations, costs and benefits, likely consequences and side effects, and a decision, on balance, that the ‘reasons for’ defeat the ‘reasons against’. Another familiar metaphor is that of weighing or balancing the relevant considerations against one another.\(^57\)

Obviously, what is reasonable depends on the context, for one can only determine which considerations are relevant and what importance to attach to each in the light of concrete circumstances.

So interpreted, however, deciding whether an act is reasonable is identical to deciding whether that act is justified, since to say that an act is justified means nothing if it does not mean that the reasons for it defeat all countervailing reasons. But public law usually distinguishes between judgments that acts are reasonable and judgments that acts are justified. The standard of reasonableness is usually thought to be more accommodating than the requirement that acts be justified, for the former leaves latitude for differences of opinion, for rival, but plausible views of what is justified in the circumstances.

Public law draws this distinction for two reasons. The first is the pervasiveness of reasonable disagreement concerning how one should act in situations of choice.\(^58\)

Frequently, reasonable people will form conflicting opinions about what kinds of considerations are relevant as well as how much weight to assign to each; in doing so they must apply concepts, including moral and political concepts, that are vague and subject to hard cases; to some extent, the way they assess the evidence and weigh the relevant moral and political values is shaped by their different life experiences; and often some of the competing factors will be incommensurable.\(^59\)

\(^57\) Although weighing and balancing metaphors capture the truth of assessing competing considerations, they are misleading to the extent that they imply that all reasons have a comparable dimension of ‘weight’ measurable in a shared metric. That is not the case, for many reasons are incommensurable. Two competing reasons, and thus the different actions they support, are incommensurable where neither reason is weightier than the other and yet they are not equal in weight either. J Raz *The morality of freedom* (1986) 322. In such circumstances, neither reason is defeated, so it is reasonable to choose either action: J Gardner & T Macklem ‘Reasons’ in J Coleman & S Shapiro (eds) *The Oxford handbook of jurisprudence and legal philosophy* (2004) 440 470 - 474.

\(^58\) These are the factors discussed by J Rawls in *Political liberalism* (2005) 56 - 57.

\(^59\) See n 57 above.
Moreover, given that the state is necessarily limited in the values it can pursue, the legislature and executive must make some selection from among the full range of political and moral values that might be realised. In the circumstances, we may expect that people will form a range of conflicting views about what should be done, many of which cannot be dismissed as unreasonable.

The second reason is the institutional position of the courts. The separation of powers, as concretised in the Constitution, envisages that the responsibility for choosing policies and initiating legislation to serve them lies with the executive, the responsibility for finalising the content of and passing legislation lies with the legislature, and the responsibility for implementing legislation lies with the public administration and other organs of state. The courts owe the decisions of these institutions particular respect, for several overlapping reasons: The judicial process, with its adversarial, bilateral procedure and fact-finding technique, is generally a poor way to make political decisions; judges tend to have limited expertise and experience in making political decisions and deciding how best to implement them; and the public’s democratic ability to remove politicians from office suggests that they, rather than judges who cannot be voted out, should be politically responsible for these decisions and should not be able to disclaim that responsibility on the ground that a decision is approved or required by judges. These overlapping reasons may be referred to compendiously as ‘institutional competence’ and ‘democratic principle’.

The pervasiveness of reasonable disagreement and the institutional respect that courts owe the political branches of government explain the distinction that public law draws between justified or correct decisions – an assessment that judges should not make – and reasonable or justifiable decisions – an assessment that judges may sometimes make. The latter standard, although fundamentally dependent on the idea that reasons for a decision outweigh the reasons against it, envisages a sphere, margin, or range of reasonable decisions within which the politically responsible persons may freely choose, by judging for themselves whether a particular decision is justified. A court assessing the reasonableness of that choice has to decide whether it falls within, or beyond, the range of reasonable decisions in the circumstances.

We also know that constitutional law draws a further distinction between its notions of rationality and reasonableness. The former, as described above, is the minimum constitutional baseline applying to all law and conduct, whereas the latter is a more stringent standard that applies

60 See, eg, the discussion in Glenister v President of South Africa 2009 1 SA 287 (CC) from para 29.
62 For this use of terminology, see W Wade & C Forsyth Administrative law (2009) 308 - 309 and Price (n 2 above) 588 - 590.
less frequently. To my mind, a profitable interpretation of that distinction is this. *Whereas a law or act is reasonable if the reasons for it defeat the reasons against it, that law or act is merely rational if, notwithstanding the reasons against it, there is at least one reason or rationale for it.*\(^6^3\) The distinction is clear: Whereas an evaluation of the reasonableness of a law or act calls for an assessment of the competing considerations for and against it, an evaluation of its constitutional rationality calls only for the identification of one rationale in its favour. In the courts’ idiom, all that is required is that the law or act be rationally connected to one legitimate purpose. To recall, that will be the case if and only if the law or act is both intended to serve a legitimate end (the purpose requirement), and does serve that end to some extent (the effect requirement), whether symbolically (where the law or act has intrinsic value) and/or by tending to bring about its end as a matter of causality (where it has instrumental value). This analysis, in turn, entails that the range of rational laws and acts in a particular context will be wider than the range of reasonable laws and acts. For many laws and acts which, on balance of the competing reasons, are unreasonable will nevertheless be rational because a legitimate rationale in their favour can be identified within the balance of competing considerations. In other words, the latitude that the courts grant to the political branches to make rational decisions is far wider than the latitude they grant to make reasonable decisions.

3.2 The principle of comity

This explanation of the difference between reasonableness and rationality has two implications. The first is that it helps us to understand the duty of courts, when applying the rationality principle, not to ‘second-guess’ the wisdom of the legislature and executive. This, it will be recalled, is convenient shorthand for various judicial statements to the effect that rationality review does not concern the political merits or demerits of the law or conduct in question, nor does it involve the making of policy choices, nor the substitution of the court’s opinions as to what is correct or appropriate for the opinions of the relevant political branch of the state.\(^6^4\) These *dicta* are, of course, a reflection of the general principle – which we may refer to as the ‘principle of comity’ – that the courts ought to respect the views of the legislature and executive, which in turn is justified by the considerations of institutional competence and democratic principle discussed above. It applies whenever the courts scrutinise law or state conduct, and requires that they adhere to their proper role within the constitutional scheme of the separation.

---

\(^6^3\) I want to emphasise that my intention, here, is to attempt to advance a clear and constructive explanation of the rationality/reasonableness distinction in South African constitutional law that is tolerably faithful to the reasoning and decisions of the courts. It is not my view that courts should jettison the current test for rationality in favour of some kind of ‘one-reason-for’ test. For an account of the nature of ‘reasons for action’, see generally Gardner & Macklem (n 57 above).

\(^6^4\) See the judgments cited above (n 52 - 56 above).
of powers.\textsuperscript{65} It is important to realise that what the principle of comity requires of courts varies with the circumstances.\textsuperscript{66} For the moment, however, we must consider what the principle of comity requires in the context of rationality review. Bishop seems to argue that the duty implies that judges should exercise no discretion at all. As he puts it:

If the test permitted or required judges to use their own discretion in determining the rationality of a choice, that command [ie the command not to second-guess legislative choices] would be undermined as judges' attitudes would pollute the so-called neutral inquiry. The traditional underpinnings of rational basis review therefore demand a test that forces predictable outcomes and removes personal discretion. Judges are therefore obliged to frame the rational basis test as excluding their discretion.\textsuperscript{67}

It seems to me that this interpretation of what the principle of comity requires in the context of rationality review is too strict for two connected reasons. First, it renders rationality review incoherent by sheer definitional stipulation. For it is beyond doubt, as Bishop convincingly argues, that the application of the rationality principle often involves the exercise of judicial discretion. This, as I have explained, is due to the evaluative nature of its purpose requirement and the vagueness of its effect requirement.\textsuperscript{68} Rationality review inevitably contradicts the duty not to 'second-guess' the decisions of the legislature and executive if that duty is interpreted strictly, as Bishop does, to exclude discretion altogether.

But, secondly, we need not interpret the duty so strictly: there is an alternative interpretation that avoids Bishop's stark conclusions. Because the rationality principle requires only that the law or conduct in question serve one acceptable purpose, it envisages a range of rational laws and acts even wider than that of reasonable laws and acts. Within that broad range of rational options, the political branches of the state are free to choose without constraint. Courts engaging in rationality review have no role to play in regard to this political choice from among rational options. Thus, on the one hand, it is a requirement of the principle of comity that judges do not 'second-guess' policy decisions within this legitimate sphere. On the other hand, however, courts are obliged to police the boundaries of that sphere by invalidating any law or conduct that lies outside it. They do so by applying the rationality principle.

Of course, on every occasion that an applicant mistakenly argues that a rational law or act is irrational, he or she effectively invites the courts to violate the principle of comity. That is because, by asking the court to invalidate rational law or conduct the applicant effectively asks the court to order the political branches not to choose an option lying within their

\textsuperscript{65} Endicott (n 61 above) 101 - 104.
\textsuperscript{66} I consider such variability below, in sec 3.3.
\textsuperscript{67} Bishop (n 5 above) 27.
\textsuperscript{68} See the discussion above, from sec 2.3.
legitimate sphere of choice. This explains why courts invariably restate their
duty not to ‘second-guess’ the political branches in those cases where
challenged laws and acts are upheld as rational. But in cases where the
challenged law or conduct fails to serve at least one acceptable purpose, the
courts should invalidate it. To the extent that they exercise discretion in doing
so, that is consistent with the principle of comity because the courts are
defining the boundaries of rational state decision making, rather than
interfering with the political branches’ freedom to choose within those
boundaries.

Of course, none of the above justifies rationality review, including the fact
that judges often exercise discretion when defining the boundaries of
rational state decision making. I consider that deeper, normative question
in parts 4 and 5 below. Instead, the advantage of this explanation is that it
renders intelligible the courts’ claim to be able both to apply the vague and
evaluative legal standard of rationality and simultaneously to pay the
executive and legislature the institutional respect they deserve.

3.3 Variability

The second implication of my account of reasonableness and rationality is
that it enables us to explain both how and why the ‘intensity’ of judicial
scrutiny of law and state conduct varies. There are two distinct forms of
variability. The first, most obvious form is between standards of review.
The rationality principle, as described above, is the minimum
constitutional standard for all law and conduct, whereas the higher
standard of reasonableness applies in a more limited range of cases: for
example, to decide whether a limitation of a fundamental right is
‘reasonable and justifiable’ in term of section 36 of the Constitution; to
decide whether Parliament has complied with its constitutional obligation

---

69 Thus, we need not accept Bishop’s suggestion in Woolman & Bilchitz (n 5 above) 27, that there is ‘a massive gap’ between what the Court does and what the Court says it does.

70 They often exercise discretion, but not always, because some cases are straightforward. Arguably, Kruger (n 30 above) and Van der Merwe (n 13 above), for example, fall into this class.

71 In this regard, ‘proportionality’ is best understood not as a standard of review entirely
different to reasonableness, but rather as a rigorously structured test of reasonableness. So
understood, a law that is disproportionately tailored to its end, and thus violates a
fundamental right, suffers from a species of unreasonableness. (The same is true of laws that
constitute unfair discrimination in terms of sec 9(3) of the Constitution on an application
of the ‘impact’ test.) It is also worth observing that rationality is a necessary condition for
proportionality. That is because, in order for a law that limits a fundamental right to be
justified in terms of sec 36, it is necessary but insufficient that it serve some legitimate
purpose (i.e., it is rational). For it is impossible for an irrational law that limits a right to be
justified under sec 36, because its cost (limiting a right) cannot possibly be outweighed by
any countervailing benefit.
to facilitate public involvement in the legislative process; to decide whether an official in the public administration has complied with the right to just administrative action and PAJA, and so forth.

But why does the rationality principle apply as a general baseline while the standard of reasonableness applies in a more restricted fashion?

Here is an answer. On the one hand, the principle of comity obliges courts to respect the decisions of the political branches of the state in all circumstances. Usually, only the relatively deferential standard of rationality is compatible with this ever present duty. On the other hand, certain laws and state conduct trigger vital individual and collective interests in a way that calls for special attention. The most obvious examples are those acts and laws that threaten or limit fundamental rights. The higher reasonableness standard is appropriate in these circumstances because the engaged interests are of such normative importance that they partially offset the courts’ countervailing duty to respect the political autonomy of the executive and legislature. This duty is never totally outweighed, though, as is reflected in the distinction the law draws between judging acts as reasonable and judging acts as justified or correct.

The second form of variability of judicial scrutiny of law and conduct is within the standards of review. This is obvious in the case of reasonableness, as different factors of variable weight must be assessed in different contexts. The greater the potential or actual adverse effects of the law or conduct in question, the stronger the required countervailing reasons militating in its favour. This variability was expressly recognised by O'Regan J in Bato Star Fishing (Pty) Ltd v Minister of Environmental Affairs.

Less well recognised, however, and more subtle is variability within rationality review. To recall, that principle requires that the law or conduct in question, once appropriately specified, first have an appropriate purpose (the purpose requirement) and, secondly, serve that purpose sufficiently well in order to be ‘rationally connected’ to it (the effect requirement). The wide variety of circumstances to which this test applies, the discretion its application involves, and the courts’ duty to apply it in a manner that respects the autonomy of the political branches of the state inevitably, and rightly, call for a degree of variability.

72 Doctors for Life International v Speaker of the National Assembly 2006 6 SA 416 (CC) paras 118 - 129, 204; Matatiele Municipality v President of South Africa 2007 6 SA 477 (CC) paras 36 - 40, 45.
73 Bato Star Fishing (n 34 above) para 44.
74 This is exactly the point that Bishop (n 5 above) questions in general. I have already explained why the claim that rationality review is compatible with the principle of comity is intelligible. I argue that that claim is justified below, from 58.
75 Note that sec 36 empowers courts to determine whether limitations to rights are ‘reasonable and justifiable’ (my emphasis), not justified. The distinction, albeit a fine one, is important.
76 Bato Star (n 34 above) para 45.
Hence, in what may roughly be labelled ‘political’ contexts, when courts assess the rationality of decisions lying close to the heart of executive power (for example those concerning foreign relations, national security, economic policy, and so forth), the court is bound to exercise considerable deference when deciding whether the purpose of the decision is legitimate and whether the decision is likely to serve that purpose. In such contexts, the considerations of democratic principle and institutional competence are particularly pressing and consequently oblige the court to pay a higher degree of respect to the empirical and normative judgments of the executive. This extra respect should be reflected in the way the court exercises its discretionary judgment, when specifying the law or conduct in question more or less widely, discerning and evaluating its purpose at a narrower or wider level of abstraction, and choosing how well the law or conduct ought to have served that purpose – how ‘closely connected’ to its end it should have been. By contrast, in less ‘political’ contexts – that is, in cases where the considerations of institutional competence and democratic principle are less weighty – the courts need not tread as lightly when evaluating the purpose and effect of the law or conduct in question. Of course, the distinction between these cases and more ‘political’ cases is one of degree and a matter of difficult, discretionary judgment for experienced judges.

3.4 Upholding the distinction between reasonableness and rationality

There is a second way in which rationality review could conceivably vary in intensity, which would lead to the distinction between rationality and reasonableness breaking down and becoming one of degree rather than kind. To recall, the rationality principle requires only a single legitimate rationale supporting the law or conduct in question, notwithstanding the presence or absence of other reasons for or against it. Such a rationale is present whenever the law or conduct serves a legitimate governmental purpose. Harmful side

---

77 Discussed above, in sec 2.3. For further elaboration of this point, see Price (n 2 above), which differentiates between ‘timid’ and ‘assertive’ instances of rationality view.

78 An example is *Van der Merwe* (n 13 above), where the Court declared irrational a statutory rule which prohibited persons married in community of property (but not persons married out of community of property) from suing their spouse in delict for ‘patrimonial damages’. The rule was an historical anomaly. At common law, spouses married in community of property originally could not sue one another in delict because they jointly owned all their property. Later, legislation came to recognise separate property falling outside the joint estate and accordingly permitted inter-spouse delictual claims, but only for non-patrimonial damages. The law thus came to permit persons married in community of property to sue their spouse only for non-patrimonial damages, but permitted persons married out of community of property to sue their spouses for both patrimonial and non-patrimonial damages. The Court unanimously held that this differentiation served no legitimate purpose, given that legislation had since recognised the possibility of spouses married in community.

79 The Constitutional Court already seems to accept this: see Poverty Alleviation Network (n 2 above) para 74, where Nkabinde J recently held for a unanimous Court that judicial intervention on the basis of rationality ‘should be guided by the principle of separation of powers’.
effects, possible alternative laws or acts, and other countervailing considerations are not relevant under rationality review. Nevertheless, one can readily imagine a court relaxing this approach by taking into account a particularly notable reason against the law or conduct, and deciding whether that reason is outweighed by any legitimate purpose that the law or conduct serves.

The unfortunate scenario facing the Constitutional Court in Merafong provides an example. One of the issues in that case was the rationality of the provisions of the Constitution Twelfth Amendment Act that purported to alter various provincial boundaries so that the Merafong City Local Municipality, which initially lay partly in Gauteng and partly in the North West province, would be located entirely within North West. It was agreed that abolishing cross-border municipalities in order to facilitate efficient local government service delivery was a legitimate purpose in general, and the majority held that the provisions in question served that purpose and therefore were rational. In doing so, however, the Court treated as irrelevant the Merafong community's burning desire, expressed with passion and occasional violence, to be located in Gauteng. That desire was clearly a powerful reason for Merafong to be put in Gauteng rather than North West, an avenue that would equally have abolished the pre-existing cross-boundary municipality. We can imagine a different court, moved by the plight of the Merafong community, deciding to take this factor into account under rationality review. In doing so, the court would have to weigh it against the purpose of locating Merafong in North West. But that would be a step in the direction of reasonableness review, for the wider the range of factors a court takes into consideration, the more closely the test comes to resemble one of reasonableness.

So one can readily imagine a spectrum: at the end of full-blown reasonableness, courts take into account all relevant considerations; at the opposite end of bare rationality, courts search only for a single legitimate rationale for the law or act, ignoring all other considerations. An approach that took into account the feelings of the Merafong people would lie somewhere in between. This illustrates how the distinction between rationality and reasonableness could become one of degree if the courts chose to depart from bare rationality by taking countervailing considerations into account.

80 n 22 above.
81 Merafong (n 22 above) paras 110 - 115.
82 That the various legislatures saw fit, at the time, to ignore the deeply-held wishes of the Merafong people is what made the case so difficult and may, from a 'legal realist' perspective, provide a plausible explanation for the minority's unprecedented willingness to declare irrational the decision of the Gauteng provincial legislature to approve the Bill in question in terms of secs 74(3)(b) and (8) of the Constitution, thereby extending rationality review deeply into the legislative process. See the discussion above (n 24 above).
The content and justification of rationality review

Yet it is clear that the courts, when applying the rationality principle to laws and executive conduct, have not expressly adopted a ‘spectrum’ approach to constitutional rationality review – according to which judges move closer to and further away from a full reasonableness standard depending on the context. Nor does it appear that they have implicitly or effectively done so. This should be welcomed. Adopting a ‘spectrum’ approach to reasonableness and rationality would make it very difficult to predict what degree of scrutiny is to apply in future cases, and would make it easier for courts to violate the principle of comity by straying beyond their appropriate constitutional role. Nevertheless, it is highly likely that applicants in rationality review cases will try to persuade judges to move in this direction, whether expressly or merely in effect. So we should be aware of both the temptation for courts to take this step, as well as the risk that they might do so implicitly or unconsciously. In my view, if the Constitutional Court ever decides that bare rationality is too deferential a standard for a particular context, it should instead openly discuss and decide whether a change to a more rigorous standard of review is justified in the circumstances.

3.5 Arbitrariness

The final issue to be briefly explored in this part is the notion of arbitrariness in the context of rationality review. Although arbitrariness is a difficult concept which has had various, shifting senses in South African law at different times and in different contexts, its meaning as employed by courts engaged in rationality review is reasonably clear. Endicott defines arbitrariness in its ‘ordinary pejorative sense’ as follows: ‘an unjustified act, a capricious or despotic act that calls – and lacks – some justification other than the fact that the actor willed it and did it.’ On this interpretation, an arbitrary act is simply an unjustified act. That cannot be what arbitrariness means in the context of rationality review in South African constitutional law. Instead, in this latter context, arbitrariness is identical to irrationality. A law or act is irrational or arbitrary simply if it fails to serve a legitimate government purpose.

---

83 I leave open the question whether the courts have adopted a ‘spectrum’ approach to rationality and reasonableness when reviewing ordinary administrative action in terms of PAJA or sec 33 of the Constitution. On this, see the sources cited above (n 34 above).

84 See, eg, the dissenting judgment of O’Regan J in New National Party (n 2 above) para 122.

85 Eg, F Michelman observes in ‘The rule of law, legality and the supremacy of the Constitution’ in S Woolman et al (eds) Constitutional law of South Africa (2005) 2nd ed 11-1 n 5, that ‘it is notorious … that the meaning of “arbitrary” underwent contraction … under the stress of apartheid-era realpolitik’. Another example is the prohibition on ‘arbitrary deprivation of property’ in terms of sec 25(1), where arbitrariness has been interpreted to lie somewhere between ‘mere rationality’ and proportionality: First National Bank v Commissioner, South African Revenue Services 2002 4 SA 768 (CC) para 65.

86 Endicott (n 61 above) 90.
Bishop argues that the rationality principle prohibits two classes of law or conduct – those motivated only by prejudice or ‘naked preferences’; and those that fail to serve an appropriate purpose – and that these two classes correspond to the twin evils that rationality review is designed to avert: prejudice and arbitrariness. It should be noted, however, that prejudice or ‘naked preferences’ are merely species of arbitrariness. So it seems more accurate to say that rationality review is designed simply to guard against the single evil of arbitrariness or irrationality in laws and conduct, and that a law or act that serves no legitimate purpose, but is instead motivated purely by prejudice, is a paradigmatic example of such arbitrariness or irrationality.

4 The justification of rationality review in principle

Armed with a clear understanding of the elements of the rationality principle, as well as of arbitrariness, reasonableness, variability of judicial scrutiny, and the principle of comity, we may now turn to our central question: whether rationality review as it applies in South African constitutional law is justified. As this is a normative question of some complexity, reasonable people are bound to differ. For instance, Bishop argues that rationality review needs to be saved, because, in his view, its malleability unmoors it from its traditional justifications and ‘prevents [it] from filling the role it is meant to serve’. He also argues:

In general, the test vacillates between being a rubber stamp for any government plan no matter how ludicrous, and a barrier to reasonable regulation with which judges disagree for other substantive reasons.

It seems to me that these criticisms attack both rationality review in principle, as well as its application by the Constitutional Court in practice. These questions may profitably be examined apart: If rationality review is unjustifiable in principle, we need not consider how courts have applied it; but even if it is justifiable in principle, the courts may nevertheless tend to apply it inappropriately. With that distinction in mind, I advance an alternative view to Bishop’s. In this part, I argue that rationality review is

87 Bishop (n 5 above) 9.
88 Bishop (n 5 above) 7.
89 Bishop (n 5 above) 25. What Bishop means by this is that ‘there is room for judicial manipulation that can drastically alter the outcomes of the test’.
90 Bishop (n 5 above) 26.
91 As above.
92 As above.
justifiable in principle.\textsuperscript{93} The core of my argument is that its limited discretionary character should be welcomed, not condemned. I also consider the courts’ frequent claim that rationality review is justified by the rule of law. In part 5, I argue that the Constitutional Court has, on the whole, applied the rationality principle in a satisfactory manner.

4.1 The enterprise of justification

Before we focus our normative attention on rationality review, we must first consider how any law could be justified. As I have already explained, a law is justified if the reasons for it defeat the reasons against it. In other words, its benefits (including its beneficial side effects) must outweigh its costs (including its harmful side effects). A legal rule or principle, generally speaking, can have two kinds of benefit. The first are its substantive benefits, the good it brings into the world by virtue of its content. That much is obvious. The second are its formal benefits, the good it brings into the world by virtue of its form. Laws that are clear, certain, prospective, stable, general and open, and upheld by independent and impartial courts, bring good into the world by reducing the risk that public power will be used for private ends and by guiding the behaviour of persons in the community, facilitating the valuable ‘ability to choose styles and forms of life, to fix long-term goals and effectively direct one's life towards them’.\textsuperscript{94}

But often the respective potential benefits of a law’s content and its form pull in different directions.\textsuperscript{95} As Du Bois observes:

There is an inescapable potential tension between [formal] legality and the pursuit of substantive objectives such as justice … The upshot is that there is a perpetual search for the right balance between restricting and empowering law-applying officials, between the values secured by the form of the law and those pursued by its substance. Unsurprisingly, the point at which this balance can be achieved is a source of constant controversy, and varies across time, between fields of law, and from one legal system to another.\textsuperscript{96}

Similarly, Hart argued:

In fact all systems, in different ways, compromise between two social needs: the need for certain rules which can, over great areas of conduct, safely be applied

\textsuperscript{93} I assume, for present purposes, that the invalidation by judges of properly passed legislation on the ground that it conflicts with a substantive requirement of a supreme, entrenched constitutional document is not unjustified in principle. That is a difficult question which lies beyond the scope of this essay. For a criticism of this assumption, see J Waldron \textit{Law and disagreement} (1999).

\textsuperscript{94} J Raz \textit{The authority of law} (2009) ch 11. So it is mistaken to describe the formal merits of laws as ‘value-neutral’ (see, eg, I Currie & J de Waal \textit{The Bill of Rights handbook} (2005) 12, referring to the formal aspects of the principle of legality), for adherence to the formal aspect of the rule of law is an important, substantive virtue of legal systems working well. It is more accurate to describe these formal merits as ‘content-neutral’.

\textsuperscript{95} See, eg, \textit{Affordable Medicines} (n 2 above) para 108 and Kruger (n 30 above) para 62.

by private individuals to themselves without fresh official guidance or weighing up of social issues, and the need to leave open, for later settlement by an informed, official choice, issues which can only be properly appreciated and settled when they arise in a concrete case.  

These insights reveal that when structuring legal rules and principles, different balances may reasonably be struck between the need for certainty and the need for discretionary, case-specific judgment. On occasion, it will be justifiable to sacrifice a degree of legal certainty, and thereby empower judges to exercise greater discretion in order to promote other substantive goals. Of course, going too far in this direction dilutes the valuable guidance that law can provide and increases the risk that the law in question will be arbitrarily applied by officials or judges. Exactly where the balance should lie will vary with the context. Open-ended legal standards, such as reasonableness, are more appropriate in circumstances ‘where the sphere to be controlled is such that it is impossible to identify a class of specific actions to be uniformly done or forborne and to make them the subject of a simple rule’, and yet some form of legal oversight is nevertheless valuable. Of course, the overall justification of such discretionary standards will also depend on the manner in which judges tend to apply them.  

4.2 Costs and benefits  

It is time to assess the justifiability of the rationality principle. First, let us consider its benefits. In Prinsloo, the majority held that the rationality principle  

has been said to promote the need for governmental action to relate to a defensible vision of the public good, as well as to enhance the coherence and dignity of legislation. In Mureinik’s celebrated formulation, the new constitutional order constitutes ‘a bridge away from a culture of authority … to a culture of justification’.  

These important truths are captured by saying that rationality review  

97 HLA Hart *The concept of law* (1997) 130.  
98 Hart (n 97 above) 132.  
99 See part 5 below.  
100 *Prinsloo* (n 2 above) para 25 (footnotes omitted). The reference is to E Mureinik ‘A bridge to where? Introducing the interim Bill of Rights’ (1994) 10 *SAJHR* 31 32.
promotes adherence to public reason.\textsuperscript{101} Public reason, simply put, is that set of principles and other reasons for action that are appropriate morally to justify laws and state conduct. On any acceptable political morality, all branches and spheres of the state bear a special role-based moral obligation always and only to pass laws and to act in a manner that is publicly justified. The rationality principle is designed to promote publicly justified state law and conduct by empowering judges to uphold, as a minimum constitutional baseline, a prohibition on arbitrary law and conduct. Of course, the legal prohibition on arbitrary or irrational law and conduct, as I have explained, is weaker than the moral prohibition on unjustified law and conduct. This, to recall, flows from the principle of comity – the institutional respect courts owe the political branches of government. Nevertheless, for state law and conduct to be publicly justified, it must at the very least be neither arbitrary nor irrational.

Rationality review promotes adherence to public reason at least in two ways. First, its application ensures that many arbitrary acts and laws are invalidated.\textsuperscript{102}

Secondly, it probably has a positive impact on political decision-making by providing the executive and legislature an incentive to ensure that their acts and laws serve legitimate purposes.\textsuperscript{103}

However, rationality review is not without its costs. Two stand out. The first is that its discretionary character injects a degree of uncertainty into the law, so that both political decision-makers and potential litigants are given imperfect guidance. The second cost is that rationality review empowers unelected judges to evaluate and, in certain restricted circumstances, to invalidate decisions of the executive and legislature. This may be costly given the limited institutional competence of courts – they are poorly suited to making political decisions, due to the nature of judicial procedure and the particular expertise of judges. Democratic principle, moreover, suggests that the political responsibility of politicians should not be diluted by overreaching judicial decisions.

\textsuperscript{101} This claim is similar, but not identical, to Bishop’s suggestion in Woolman & Bilchitz (n 5 above) 29, following Mureinik (n 100 above), that rationality review promotes ‘a culture of justification’. A culture of justification, in a political community, is a predominant practice of seeking to persuade the public that laws and state conduct are justified; genuine state adherence to public reason, on the other hand, requires those justifications to be legitimate and convincing. In my view, while both goals are extremely valuable, the latter is more important and fundamental. I believe that rationality review serves both.

\textsuperscript{102} Rationality review cannot ensure that all state arbitrariness is legally invalidated. Given the huge role that the state plays in South African society and the sheer number of its laws and acts, as well as the high costs of litigation, it is possible that many irrational laws and acts slip through the net.

\textsuperscript{103} I say ‘probably’ because this is essentially a contingent matter: Empirical research would be required to be certain about any concrete impact that court orders might have on political decision making.
Do these costs outweigh rationality review’s benefits, thereby rendering it unjustified in principle? I think not, for there are at least three reasons why we should not exaggerate the costs. First, because the rationality principle applies across such a broad range of contexts, it has to be structured in an open-ended way that calls for case-specific judgments. It would be impossible for courts to guard against arbitrariness in legislation and the exercise of public power by applying a more rigid rule. For this reason, the flexible, discretionary character of rationality review is part of its value. This suggests that it is mistaken to claim that ‘[t]he traditional underpinnings of rational basis review ... demand a test that forces predictable outcomes and removes personal discretion’.\(^{104}\) Secondly, we should also not over-exaggerate the uncertainty of rationality review. The law instructs all political decision makers to act and legislate in a manner that serves a legitimate purpose. That, in itself, is significant guidance. In addition, although it is possible that different judges may exercise discretion under rationality review inconsistently, the risk of contradictory decisions remaining in the law reports is reduced by the fact that only the Constitutional Court can finally invalidate law or the conduct of the President.\(^{105}\) Furthermore, the respect the Court owes the political branches under the principle of comity suggests that laws and conduct should be declared invalid only rarely.\(^{106}\) Finally, as my account of the duty not to ‘second-guess’ the decisions of the executive and legislature makes clear,\(^{107}\) courts are not permitted to evaluate and approve or disapprove of political decision-making in an entirely open-ended way; instead, they must sometimes exercise a constrained discretion to guard against an egregious, and therefore hopefully rare, evil – state arbitrariness or irrationality.

For all these reasons, it is far from obvious that rationality review is unjustified in principle. The promotion of public reason could hardly be more valuable, and empowering courts to guard against irrational laws and acts contributes to this end. Introducing a degree of legal uncertainty and empowering judges to evaluate political choices in order to draw the outer boundaries of legitimate state decision making is arguably a fair price to pay. Imperfect legal guidance and a limited invasion of the state’s political autonomy are not obviously disproportionate to the moral gains on offer, in the form of promoting adherence to public reason.

4.3 The rule of law as a justification

The final matter to be addressed in this part is the courts’ frequent claim

\(^{104}\) Bishop (n 5 above) 9.
\(^{105}\) Sec 167(5) of the Constitution.
\(^{106}\) See the discussion in part 5 below.
\(^{107}\) Sec 3.2.
that rationality review is justified by the rule of law, which is a foundational constitutional value of great importance. The argument seems to be that, because arbitrary governance is by its nature inconsistent with the rule of law and the idea of a constitutional state, it follows that the courts are justified in upholding a prohibition on arbitrariness by assessing the purposes and effects of laws and state conduct. What can we make of this argument?

For one thing, if it is sound, then it gives rise to a puzzle. On the one hand, it should by now be clear that the rationality principle both suffers from a degree of uncertainty and affords the Constitutional Court a discretion that may be misapplied or abused, from which there is no further appeal or review. In these two respects, rationality review appears to undermine some of the requirements of the rule of law. But, on the other hand, it is the rule of law itself that is said to justify rationality review in the first place. How can it be that the rule of law justifies the rationality principle if its application in fact undermines the rule of law?

The way to resolve this paradox is to recognise that the rule of law, as a justiciable legal principle in South African law, has been held by the courts to impose restrictions on both the form and the content (or ‘substance’) of laws and conduct. For example, the requirement of ‘reasonable certainty’ is a formal requirement, while the requirement of rationality or non-arbitrariness is a requirement of content. Although the test for arbitrariness is not perfectly clear, the resulting deficit in the formal aspect of the rule of law is arguably offset by the countervailing substantive gains arising from the invalidation of arbitrary laws and acts.

The Constitutional Court thus seeks to justify rationality review by appealing to a ‘substantive conception’ of the rule of law, which, as we have seen, imposes various requirements of form and content that do not always pull in the same direction. It follows that the courts’ claim that rationality review is justified by the rule of law necessarily depends, in turn,

---

108 See, eg, the cases cited (n 4 above). Reference is also made in Makwanyane (n 2 above) para 156 and Affordable Medicines (n 2 above) para 74 to ‘the idea of the constitutional state’.

109 These, of course, are the formal requirements of the rule of law. See generally Raz (n 94 above), as well as Affordable Medicines (n 2 above) para 108 and Kruger (n 30 above) para 64.

110 There is an important distinction between the rule of law as a justiciable principle of law, which is given concrete meaning by judicial decisions, and the less determinate, political value of the rule of law, which forms part of political morality. Accordingly, one could coherently accept that Pharmaceutical Manufacturers (n 2 above) held that the rule of law, as a justiciable legal principle, imposes a substantive requirement of rationality that courts must uphold and enforce, but also argues that in doing so the Court misunderstood the value of the rule of law as a part of political morality.

111 Affordable Medicines (n 2 above) para 85; New National Party (n 2 above) para 24 and United Democratic Movement (n 2 above) para 55.

on a prior sound argument justifying their adoption of the substantive conception. I shall not consider here whether such an argument exists, for that is a difficult question and the well-worn debate about the nature, justification and implications of the rule of law continues unabated. For this reason, the courts’ appeal to the rule of law as a justification for their applying the rationality principle is perhaps somewhat controversial. Moreover, in the light of the alternative justification for rationality review advanced above, that appeal is arguably unnecessary. If we can accept that rationality review is justifiable as a means to promote adherence to public reason, because it reduces the incidence of arbitrary laws and state acts, we need not appeal to a substantive conception of the rule of law.

5 The justification of rationality review in practice

If, as I have argued, we have reason to believe that rationality review is defensible in principle, then our normative gaze must shift to the manner in which it has been applied by the courts. Judges wield a significant degree of power when they draw the boundaries of legitimate political decision-making by applying the rationality principle, and like all power, it can be misused. My argument, however, will be that the Constitutional Court has, on the whole, applied the rationality principle in a sensible manner. I do not share Bishop’s conviction that rationality review has vacillated between rubber-stamping ‘ludicrous’ laws and conduct, on the one hand, and obstructing defensible laws and conduct, on the other. What follows is an overview and thematic survey of the Constitutional Court’s record thus far, and then a closer discussion of the judgments of that Court in three difficult cases.

5.1 Overview

To date, the Constitutional Court has upheld only five rationality challenges

114 See, eg, Craig (n 113 above); J Raz ‘The politics of the rule of law’ in J Raz Ethics in the public domain (1993) 370; D Dyzenhaus (ed) Recrafting the rule of law (1999); BZ Tamanaha On the rule of law (2004); N MacCormick Rhetoric and the rule of law (2005); and T Bingham The rule of law (2010).

115 It is worth repeating here that my argument is limited to applications of the constitutional principle of rationality by the Constitutional Court. I do not undertake the wider study of High Court and Supreme Court of Appeal decisions that a full assessment of rationality review under PAJA and sec 33 of the Constitution would require.

116 Bishop (n 5 above) 26. Similarly (28) he argues that rationality review ‘is either so weak that it fails to invalidate the laws it is meant to invalidate or so uncertain that it treads on the “negative” requirement underlying its existence: Courts must not interfere with legislative wisdom.’
to law or executive conduct;\(^{117}\) it has dismissed far more.\(^{118}\) Of the former set of cases, two (Pharmaceutical Manufacturers and Kruger) involved good faith mistakes by the President that were obviously irrational, while a third (Van der Merwe) involved an outdated and patriarchal statutory restriction on delictual claims by certain married persons against their spouses, which manifestly served no purpose in the light of subsequent legislative developments.\(^{119}\) A fourth (Albutt) concerned the President’s refusal to afford a hearing to victims of crimes committed for allegedly political motives when deciding whether to pardon the perpetrators under a special scheme intended to promote national unity and reconciliation. Although the decision undoubtedly extends rationality review into the realm of procedure, it is difficult to disagree with the Court’s ruling that it was irrational for the President to attempt to promote reconciliation by excluding victims from the process. The fifth and final decision (Nyathi) was admittedly problematic and I return to discuss it below. Nevertheless, the broad pattern of decisions shows that the Constitutional Court has not tended to abuse or misapply its discretionary power under rationality review so as to interfere unduly with the political autonomy of the executive and legislature. It cannot be said that the Court has, on the whole, unjustifiably ‘second-guessed’ legislative and executive wisdom.

It follows that, if the Court is guilty of any general sin, it could only be excessive deference. But in which cases has the Court rubber-stamped plainly irrational acts? When has it upheld obviously arbitrary laws? To my mind, after more than a dozen instances of rationality review, only two decisions bear any possibility of fitting this description – Union of Refugee Women\(^{120}\) and Merafong\(^{121}\) – and I defend the application of the rationality principle in both below. But even if the Court has occasionally erred in this manner, has it tended to do so? I do not think so. To substantiate this claim, let us consider how the Constitutional Court has applied the various elements of the rationality principle a little more closely.

---

\(^{117}\) Pharmaceutical Manufacturers (n 2 above); Van der Merwe (n 13 above); Kruger (n 30 above); Nyathi (n 13 above), Albutt (n 2 above) (discussed in Price (n 2 above)). I have already explained (n 13 above), why Ntuli is not a genuine example of rationality review. A possible sixth example is Zealand v Minister for Justice and Constitutional Development 2008 4 SA 458 (CC) para 44, but in that case the Court’s review for rationality is merely an obiter dictum.

\(^{118}\) Prominent examples include S v Rens 1996 1 SA 1218 (CC); Prinsloo (n 2 above); Harkson (n 13 above); Walker (n 4 above); Jooste (n 13 above); New National Party (n 2 above); Bel Porto (n 13 above); Merafong (n 22 above); United Democratic Movement (n 2 above); Union of Refugee Women (n 13 above); Weare (n 13 above); Poverty Alleviation Network (n 2 above); Law Society of South Africa (n 2 above); Glenister (n 2 above). Another example is arguably provided by Geuking v President of South Africa 2003 3 SA 34 (CC) paras 25 - 30, despite the fact that the language of ‘rationality’ was not used.

\(^{119}\) This case is discussed in n 78 above.

\(^{120}\) n 13 above.

\(^{121}\) Poverty Alleviation Network (n 2 above) stands or falls with Merafong, because the former decision rejected a rationality challenge to legislation very similar to that upheld in the latter case, and did so for the same reasons. For further discussion, see Price (n 2 above).
First, it is fairly clear that the Court has generally been sensitive to the need to specify the challenged law or conduct in an appropriate way, taking into account the broader legislative and policy context. For example, in *Prinsloo*, the presumption of negligence of neighbouring landowners in respect of fires starting on their properties was assessed in the light of the legislative scheme as a whole.\(^\text{122}\) The same was true in *Jooste*, where a legislative bar on ordinary delictual claims by employees against employers was assessed in the light of the statutory compensation scheme that was designed to replace such claims.\(^\text{123}\) So too, in *Bel Porto*, where a policy decision by the Western Cape Education Department not to take over the employment of certain persons employed directly by state schools was assessed in light of the Department’s broader policy of promoting equity in school staffing throughout the province.\(^\text{124}\)

Secondly, there is no reason to believe that the Court has generally applied the purpose requirement in an inappropriate way. On the one hand, it has been sensitive to the need to specify the purpose of the law or act at a meaningful level of abstraction. Thus, in *Van der Merwe*, Moseneke DCJ held:

>A court remains obliged to identify and examine the specific government object sought to be achieved by the impugned rule of law or provision. In other words, we are obliged to look at the specific purpose of section 18(b) [of the Matrimonial Property Act] even though the general purpose of regulating property arrangements in marriage may not in itself be open to constitutional doubt.\(^\text{125}\)

In *Weare*, for example, the Court upheld as rational provisions in an old provincial ordinance that prohibited juristic persons from holding bookmaker licences, not on the meaningless ground that they merely ‘regulated bookmaking’,\(^\text{126}\) but rather on the ground that they eased the burden of bookmaking regulation, because holding natural persons accountable for bookmaking activities was less resource intensive than holding juristic persons so accountable.\(^\text{127}\)

On the other hand, the Court has, on the whole, exercised its discretion to evaluate the legitimacy of the purposes of challenged laws and acts in a defensible manner. That is because the proffered purpose in most cases has been obviously legitimate: for example, to prevent meritless appeals from progressing in the courts;\(^\text{128}\) to incentivise fire-prevention behaviour;\(^\text{129}\) to ensure that trustees acquire all the assets of an

\(^{122}\) n 2 above, paras 39 - 40.  
^{123}\) n 13 above, para 17.  
^{124}\) n 13 above, paras 43 - 44.  
^{125}\) n 13 above, para 33.  
^{126}\) Bishop’s discussion of this case in Woolman & Bilchitz (n 5 above) from 15 - 16.  
^{127}\) n 13 above, para 50.  
^{128}\) *Rens* (n 119 above).  
^{129}\) *Prinsloo* (n 2 above).
insolvent spouse; to exercise financial prudence while pursuing equity in state school staffing during a difficult transitional period; to take temporary measures to phase in equitable municipal rates collection, while also ensuring continuity, in the light of lingering historical disparities in service delivery; to promote efficient local government service delivery by the abolition of cross-border municipalities; to promote national unity and reconciliation; and so forth. Presumably because the application of the 'legitimate government purpose' requirement has seldom proved contentious, the Court has not considered its meaning in any depth. That will remain the case if, as I have argued above, the Court adopts an approach according to which the purpose of the law or conduct in question is 'legitimate', under the rationality principle, if and only if it is consistent with all other legal and constitutional constraints, including the Bill of Rights and the Constitution's objective, normative value system.

The Court has also made it clear that only one legitimate purpose need be served for a law or act to be rational, and consequently it has not become embroiled in evaluating the many, sometimes conflicting, purposes that laws and conduct may serve. In Jooste, for example, where legislation barred ordinary delictual claims by employees against employers, the Court restricted itself to assessing whether the balance struck by the new statutory compensation scheme between the various conflicting interests of employees and employers, and thus multiple possible purposes, was an arbitrary one.

Thirdly, there is little reason to believe that the Court has generally applied the rational 'connection' or 'relationship' requirement inappropriately. This vague condition, to recall, requires that the law or act serves its purpose in some way, but specifies neither how, nor how well, it should do so. Undoubtedly, the Constitutional Court has taken a fairly deferential attitude in this respect. First, it was held that the law or conduct need not be the least restrictive means to achieve its legitimate purpose. Then, in Weare it was unanimously held:

The question is not whether the government could have achieved its purpose in a manner that the court feels is better or more effective or more closely connected to that purpose.

---

130 Harken (n 13 above).
131 Bel Porto (n 13 above).
132 Walker (n 4 above).
133 Merafong (n 22 above); Poverty Alleviation Network (n 2 above).
134 Albutt (n 2 above).
135 See sec 2.3 above.
136 See the cases cited in n 42 above.
137 n 13 above.
138 Prinsloo (n 2 above) para 35.
139 n 13 above, para 46. See also East Zulu Motors (Pty) Ltd v Empangeni/Ngwelezane Transitional Local Council 1998 2 SA 61 (CC) para 24.
Thus, it seems that the Constitutional Court will usually uphold as rational laws and conduct that somehow serve their legitimate purposes. This degree of deference is understandable, in general, given the institutional respect courts owe the political branches of the state. It follows that if the law or conduct in question fails to serve its legitimate purpose at all (as in Kruger and Pharmaceutical Manufacturers), it will be declared arbitrary; conversely, it will be upheld as rational if it either achieves its end symbolically or tends to achieve its end as a matter of causality. 140 Most decisions have fallen into the latter category, where the law or conduct is upheld because the Court accepts as an empirical fact that it tends, or will tend, to serve its purpose. For example, in Prinsloo the Court approved the empirical assumption that the presumption of negligence of landowners outside fire-controlled areas in respect of fires spreading from their land would incentivise vigilance against fires on their part. 141 In Merafong, it was accepted that the abolition of cross-boundary municipalities would lead to more efficient state service delivery. 142 In Walker, the Court held that the Pretoria City Council’s temporary policy of charging residents of Mamelodi and Atteridgeville a flat rate for electricity and water, while charging residents of old Pretoria for their actual consumption using meters, would tend to achieve equality of facilities and resources between residents of different areas. 143

Where, as in these cases, the connection between law or conduct and its purpose is a causal matter, there is inevitably a degree of empirical uncertainty. Will the means serve the end? How efficiently will it do so? Will the end be fully or merely partially achieved? And so on. Sometimes no evidence settling these questions of probabilistic causation will be available, and acquiring it (when that is possible) will often be very costly. In such circumstances, and provided no vital interests (such as those protected by fundamental rights) are threatened or limited, it is surely justified for the Court to defer to the plausible empirical assumptions of the executive and legislature. Where the state’s empirical claims seem implausible or far-fetched, however, the Court should require some evidentiary support for them. And where such evidence does exist, the Court should of course be willing to take it into account. 144 Nonetheless, in no rationality review case, in my view, has the Constitutional Court approved an implausible causal assumption, nor has it ever accepted the state’s causal assumptions in the face of contrary empirical evidence.

This short, thematic survey suggests that, generally, the practical application of the rationality principle by the Constitutional Court has not

140 This distinction is discussed above, sec 2.3.
141 n 2 above, para 40.
142 n 22 above, para 114.
143 n 4 above, para 27.
144 Bishop’s suggestion to this effect in Woolman & Bilchitz (n 5 above) 34 is surely justified.
been problematic. Despite the discretionary character of the test, the outcomes and reasoning in most cases have been defensible.

5.2 Three difficult cases

Three instances of rationality review by the Constitutional Court deserve closer scrutiny. All three cases were genuinely difficult, albeit for differing reasons, so it is not surprising that the Court was divided on each occasion. I shall supplement my defence of the Constitutional Court’s record by defending the majority’s application of the rationality principle in the first and second cases, and by offering an interpretation of the majority judgment in the third case.

We start with the judgment in Merafong145 where, to recall, the Merafong community challenged, inter alia, the rationality of legislation amending provincial boundaries to relocate the Merafong municipality, which had previously straddled the Gauteng-North West border, wholly within North West. The community passionately wanted to be placed in Gauteng. Everyone agreed that abolishing cross-boundary municipalities would serve the valuable purpose of promoting effective state service delivery. But what exactly had to be rational? Was it the decision to abolish cross-boundary municipalities by relocating Merafong in either Gauteng or North West? Or the decision to locate Merafong in North West instead of Gauteng? The choice had a direct bearing on the outcome: the former characterisation automatically ensured that the law served its purpose and thus was rational; the latter presumably called for an additional purpose to be served specifically by choosing North West over Gauteng, but since the state had not proffered one, and the relative consequences of choosing either North West or Gauteng were contested and uncertain, the law may have been arbitrary.146 This choice of characterisation is a particularly clear example of the discretion judges must sometimes exercise under the rationality principle.

The majority chose the former characterisation and upheld the law.147 In my view, this was the correct outcome, for two connected reasons. First, even under the latter characterisation, the law could not justifiably be regarded as arbitrary or irrational. Locating Merafong within North West still served the purpose of abolishing a cross-boundary municipality thereby easing service delivery problems. It could hardly be said in reply that defining the purpose at this level of abstraction renders rationality review meaningless. But secondly, as Van der Westhuizen J observed,148 adopting the narrower characterisation and investigating whether choosing North West over Gauteng served legitimate purposes over and above that of

---

145 n 22 above. Other aspects of the case are discussed above (n 24 above) and 56.
146 This is what I understand Bishop (n 5 above) to imply at 24. He observes that Sachs J had previously suggested in an obiter dictum in a minority concurring judgment in a similar case (Matatiele Municipality v President of South Africa (1) 2006 5 SA 47 (CC) paras 101-108) that the latter characterisation of the issue was appropriate.
147 n 22 above, para 114. The minority did not address the issue.
148 n 22 above, para 114.
abolishing a cross-boundary municipality would inevitably have entangled the Court in a complex political and economic debate, which, for reasons of institutional competence and democratic principle, judges should avoid having to resolve.

That is surely correct. The Court was not well-equipped to evaluate and choose between the competing economic and political arguments advanced for and against choosing North West over Gauteng.\(^{149}\) In addition, it is the elected political representatives, not judges, who are primarily responsible for making these choices, and who stand to lose their jobs if they systematically ignore the will of the people. Judicial intervention, beyond upholding a baseline standard of rationality, would have diluted that responsibility, thereby violating the principle of comity. For these reasons, the majority cannot fairly be said to have unduly rubber-stamped an obviously arbitrary law. Instead, the case provides a good example of how the considerations underlying the principle of comity do, and should, guide the way the courts exercise their discretion under the rationality principle.\(^{150}\)

The second difficult case was *Union of Refugee Women v Director, Private Security Industry Regulatory Authority*,\(^{151}\) which dealt with legislation prescribing various requirements for registration as a private security guard,\(^{152}\) one of which was that the applicant had to be either a South African citizen or a permanent resident. The section also provided that notwithstanding these requirements, ‘any applicant’ could be registered ‘on good cause shown and on grounds which are not in conflict with the purpose of this Act’.\(^{153}\) A group of lawful refugees challenged the permanent residency/citizenship condition on the ground that it was unfairly discriminatory in terms of section 9(3) of the Constitution and irrational in terms of section 9(1). The majority rejected both challenges and upheld the legislation; the minority held that the condition violated section 9(3), but some of its remarks seem to imply that section 9(1) was also violated. The following analysis relates only to the rationality question: Was the permanent residency/citizenship condition rationally related to a legitimate governmental purpose?

Bishop argues that the majority’s answer to this question ‘renders rationality review meaningless’ because it specified the law’s purpose at a

\(^{149}\) Both parties had advanced various, competing arguments for and against choosing either one of the provinces, dealing *inter alia* with service delivery, general economic well-being, emotional attachment, etc: *Merafong* (n 22 above) paras 110 - 113.

\(^{150}\) This point was accepted in *Poverty Alleviation Network* (n 2 above) para 74, where Nkabide J held for a unanimous Court that the application of rationality review ‘should be guided by the principle of separation of powers’. Happily, it appears that the wheels of democracy eventually turned: It was reported on the *Mail & Guardian* website that some months after judgment was handed down the law was again amended by Parliament to relocate Merafong from North West to Gauteng, in line with the wishes of the community: ‘Merafong to be returned to Gauteng’ *Mail & Guardian Online* http://www.mg.co.za/article/2009-04-09-merafong-to-be-returned-gauteng (accessed 12 April 2011).

\(^{151}\) n 13 above.

\(^{152}\) Sec 23(1) of the Private Security Industry Regulation Act 56 of 2001.

\(^{153}\) Sec 23(6).
meaninglessly abstract level equivalent to ‘public safety’. He also argues that the majority failed to assess the challenged provisions in the light of their legislative context, and that had it done so it would have declared them irrational. If these criticisms hit their mark, then *Union of Refugee Women* would indeed be a good example of excessive judicial deference.

But it is not clear that they do. First, the majority did not in fact specify the governmental purpose at a meaningless level of abstraction. Instead, they reasoned as follows: The overall purpose of the legislation was to regulate the private security industry in a way that protects public safety. The specific purpose of the condition that people first be registered was to ensure that only trustworthy people work as security guards, for untrustworthy guards pose a risk to safety. Thus, all the requirements for registration were designed to limit eligibility to persons whose trustworthiness could be objectively verified. The same was true of the citizenship/permanent residence requirement: for, although foreign nationals, including refugees, are not inherently less trustworthy than South Africans, the ‘unfortunate reality’ is that it is generally more difficult, as a practical matter, for such persons to verify their trustworthiness objectively.

While reasonable people may disagree about the cogency of this reasoning (especially the final empirical assumption), it is certainly not meaninglessly abstract.

Secondly, it is true that the majority failed to place the requirement in its broader legislative context under the section 9(1) enquiry. But it is not clear that, had they done so, they should have declared the law arbitrary. For the statute also empowered the relevant authority to register ‘any applicant’, including foreign persons, ‘on good cause shown and on grounds which are not in conflict with the purpose of this Act’. Thus, viewed holistically, the legislative scheme comprised, on the one hand, a blanket prohibition on registering foreign nationals, premised on the general practical difficulties in objectively verifying the trustworthiness of such applicants; on the other hand, the authority was able to register those foreign nationals who were in fact able to prove their trustworthiness. Such a scheme is surely rationally related to the purpose of ensuring that only trustworthy people, including foreigners, are registered. And surely requiring

---

154 n 5 above, 19.
155 Bishop (n 5 above) 22.
156 n 13 above, para 37.
157 *Union of Refugee Women* (n 13 above) para 38.
158 *Union of Refugee Women* (n 13 above) para 39.
159 *Union of Refugee Women* (n 13 above) para 38.
160 Bishop suggests the contrary in Woolman & Bilchitz (n 5 above) 22, adopting the minority’s reasoning at paras 118 - 119.
161 n 153 above.
foreign nationals to apply independently is rationally related to the purpose, given the practical difficulties faced, of assessing their trustworthiness on a more flexible, case-by-case basis. For these reasons, I think the majority was correct to hold that the legislation in question was not arbitrary and thus did not violate section 9(1).  

The third case is *Nyathi v Gauteng MEC of Health*, 163 where the majority held that the legislative prohibition on executing judgment debts against state assets 164 was unconstitutional on several grounds. One of those was its finding that the prohibition violated section 9(1) of the Constitution. What is striking about the majority’s application of the right to equality before the law is that the notions of rationality or arbitrariness are never mentioned. Unlike the minority judgment, 165 the majority makes no attempt to apply the rationality principle. Nor does it explain why some other approach to section 9(1) is appropriate in the context of the case. This is hard to understand. For, in my view, the prohibition on the execution of judgment debts against state assets (or, in equality terms, the differentiation between judgment creditors against the state and judgment creditors against private debtors) manifestly serves the worthwhile purpose of preventing disruption to state service delivery, including vital services such as policing, healthcare, and so forth. Whether it is the least restrictive means to that end is irrelevant under rationality review. 166 To describe this law as irrational or arbitrary is to stretch language past its breaking point.

The only reasonable interpretation of the majority judgment, then, is that it implicitly developed section 9(1) to require something more than mere rationality in the circumstances of the case. The majority hints as much by holding that the law in question ‘makes an unjustifiable differentiation’. 167 Such a development in the law may, in fact, be welcome, 168 but it is a pity that it was not discussed openly. It also follows, of course, that *Nyathi* is not an example of the misapplication of rationality review; it is a puzzling example of the implicit abandonment of rationality review.

---

162 This is to say nothing about whether the citizenship/permanent residence condition was unfairly discriminatory. It is possible for rational laws and acts to discriminate unfairly if, despite their rationally serving a legitimate purpose, they nevertheless severely offend the dignity of an identifiable class of persons, or otherwise affect them in a comparably serious manner.

163 n 13 above.

164 Sec 3 of the State Liability Act 20 of 1957.

165 The minority applies the rationality principle at paras 139 - 143.

166 Prinsloo (n 2 above) para 35.

167 Prinsloo (n 2 above) para 39 (my emphasis).

6 Conclusion

So after a long march through both theory and practice, we arrive at our conclusion. Rationality review does not need to be saved. In principle, its benefits outweigh its costs. Both the imperfect guidance it provides and the limited discretionary power wielded by the judges who apply it are arguably more than compensated for by the valuable part it plays in promoting state adherence to public reason. We need not appeal to a substantive conception of the rule of law, as the courts have done, in order to reach this conclusion. In addition, with the possible exception of the decision in Nyathi, rationality review has not been applied by the Constitutional Court in an inappropriate way.

If any generalisations can safely be made, they are these: First, the Constitutional Court has expressed a willingness to subject all law and every exercise of public power, no matter how politically sensitive, to the scrutiny of rationality review. Secondly, in circumstances where no fundamental rights are threatened or limited, the Court is reluctant to invalidate laws or acts unless they are clearly arbitrary. This, on the whole, is a well-balanced approach. On the one hand, it reminds the executive and legislature that no exercise of public power is beyond constitutional control and that every law and act should serve a purpose that is constitutionally acceptable. On the other hand, that judicial control is exercised in a restrained manner, consistent with the courts’ duty to respect the political autonomy of the other branches of government.

If there is no need to save rationality review, it follows that we need not provide it with a new structure. The approach developed over the past fifteen years by the Constitutional Court is, by and large, a defensible one, and may be followed in good conscience in the future.\footnote{Bishop (n 5 above) 32 - 33, where an elaborate new structure is proposed. In fact, in some respects his proposals overlap with the approach already developed by the Court. For example, while he proposes that, when assessing the rationality of differentiations in terms of sec 9(1) of the Constitution, the burden of justification should lie on the state, in practice it is already for the state to proffer justificatory purposes: arbitrariness is present where ‘there is no rational relationship between the differentiation in question and the governmental purpose which is proffered to validate it’ (Prinsloo (n 2 above) para 26; Harksen (n 13 above) para 44 (my emphasis)). In addition, while Bishop suggests that the challenged law should be understood in the light of its context, in practice that is what is usually done. If, however, we reinterpret Bishop’s argument as, at bottom, a call merely for greater transparency in the Court’s evaluative reasoning when applying rationality review, then there is little reason to disagree. Judicial transparency is valuable, and more of it is usually better than less (unless it comes at the expense of other judicial virtues, or results in political attacks which undermine the Court’s institutional security). Nevertheless, it would be an exaggeration to say that the Court has generally applied the rationality principle with insufficient transparency.}
TAKING DIVERSITY SERIOUSLY: RELIGIOUS ASSOCIATIONS AND WORK-RELATED DISCRIMINATION

Patrick Lenta*

1 Introduction

In a series of decisions implicating the right to freedom of religion, the Constitutional Court has indicated its readiness in appropriate cases to grant exemptions from facially neutral laws or regulations of general application where such laws or regulations impose disproportionate burdens on members of religious groups. In a recent decision, MEC for Education, KwaZulu-Natal and Others v Pillay, the Constitutional Court granted an exemption from a regulation to permit a member of a religious group to engage in a practice that (the Court determined) expressed her religious beliefs and culture. Should courts likewise be prepared to grant an exemption from anti-discrimination legislation to permit religious associations to engage in work-related discrimination on legally-prohibited grounds?

In Pillay, a central reason for the Court's decision to accommodate the pupil concerned was that the school's interest in enforcing the school uniform regulations without exception was weak. As Langa CJ observed, it was difficult to see how granting an exemption to permit the wearing of a nose stud by a pupil would interfere with the effective running of the

* Associate Professor, School of Philosophy and Ethics, University of KwaZulu-Natal. This paper was previously published in (2009) 126 SALJ 827.
school and the purposes the school uniform was designed to further. The government's interest in enforcing anti-discrimination legislation without exception appears far stronger. A claim for an exemption from anti-discrimination legislation seems in one respect more closely to resemble the claim in Christian Education South Africa v Minister of Justice than Pillay: If an exemption were to be granted, individuals' fundamental rights would be infringed. In Christian Education, the right of pupils not to suffer physical violence at the hands of teachers was held to trump the wish of teachers in a religious institution to inflict corporal punishment in accordance with their religious convictions. Since the purpose behind legislation outlawing unfair discrimination is to protect all persons' rights to equality (a fundamental right), an exemption from such legislation would seem to permit the violation of the right of individuals to be free from unfair discrimination. Surely the right of individuals not to suffer work-related discrimination on prohibited grounds should trump the interest of religious associations in discriminating in accordance with their religious beliefs?

No, not always. Certain instances in which religious associations engage in work-related discrimination on grounds forbidden to commercial enterprises attract relatively little opposition. In South Africa, anti-discrimination legislation proscribes work-related discrimination on prohibited grounds that include race, gender, sex, sexual orientation and religion. Yet, in at least one context it is clear that churches should be permitted to discriminate: Churches should be allowed to apply religious criteria in their choice of candidates for appointment to sacerdotal office. No one whose opinion is worth a moment's consideration would insist that a religious community could not make being counted on the prevailing criteria as a member of that community a condition of being appointed as a religious or spiritual leader, even if this constitutes discrimination on the grounds of religion.

---

3 Pillay (n 1 above) para 101.
4 Christian Education (n 1 above).
5 The right to equality is protected in sec 9 of the Constitution of the Republic of South Africa, 1996. Sec 9(3) of the Constitution provides: 'The state may not unfairly discriminate directly or indirectly against anyone on one or more grounds, including race, gender, sex, pregnancy, marital status, ethnic or social origin, colour, sexual orientation, age, disability, religion, conscience, belief, culture, language and birth.' Sec 9(4) reads: 'No person may unfairly discriminate directly or indirectly against anyone on one or more grounds in terms of subsection (3). National legislation must be enacted to prevent or prohibit unfair discrimination.'
6 Sec 6(1) of the Employment Equity Act 55 of 1998 reads: 'No person may unfairly discriminate, either directly or indirectly, against an employee, in any employment policy or practice, on one or more grounds, including race, gender, sex, pregnancy, marital status, family responsibility, ethnic or social origin, colour, sexual orientation, age, disability, religion, HIV status, conscience, belief, political opinion, culture, language or birth.' The Promotion of Equality and Prevention of Unfair Discrimination Act 4 of 2000 forbids unfair discrimination on prohibited grounds listed in sec 1(1)(xiiii)(a), which are ‘race, gender, sex, pregnancy, marital status, ethnic or social origin, colour, sexual orientation, age, disability, religion, conscience, belief, culture, language and birth'.
The right of churches to engage in work-related discrimination on prohibited grounds does not stop there, however. Although many regret the exclusion of women from the Catholic priesthood and the rabbinate of orthodox Judaism, most people who have given thought to the matter agree that anti-discrimination legislation should not be invoked to end this gender discrimination. The reason why anti-discrimination legislation should not be applied to prevent such discrimination is that religious associations have a right to manage their own affairs in accordance with the beliefs of members as expressed through their churches – freedom of religion and associational autonomy demands as much. Liberals can admit this (second) exception without being committed to considering the position on gender roles contained in the tenets of these churches to be morally benign, or otherwise endorsing it.

The two examples considered above relate to positions of religious leadership. Should the right of churches to discriminate in these contexts be extended to include other categories of employee and independent contractor? In a recent decision, Strydom v Nederduitse Gereformeerde Gemeente Moreleta Park, the Equality Court had to determine whether a religious association should be granted an exemption from the Promotion of Equality and Prevention of Unfair Discrimination Act (Equality Act). W Galston Liberal pluralism: The implications of value pluralism for political theory and practice (2002) 111. Roman Catholics, Muslims, orthodox Jews, Hindus and Sikhs may all claim that, according to their beliefs, the valid performance of the sacerdotal function requires that it be performed by a man. A church’s right to engage in work-related discrimination on the basis of gender (or sexual orientation, in the case of churches whose tenets require religious leaders to be heterosexual) when it comes to the ministry is, as the Colorado Supreme Court has observed, an ecclesiastical matter; a ‘church’s decision of whom to hire as a minister necessarily involves religious doctrine’ (Van Osdol v Vogt 908 P 2d 1122 (Colo 1996) 1128). Eg, a claim by a woman asking a court to intervene on her behalf because she wishes to be ordained as a priest by the Catholic Church, which refuses to consider her for the ministry, places her in a paradoxical position: on the one hand she claims to be a Catholic, which entails acceptance of the authority of the Pope; on the other she is asking a court to overrule the Pope’s stance on woman priests. For a court to rule on this matter would involve it deciding in favour of one side and against another in a wholly internal dispute within the church, thereby altering the character of the religion concerned. As Barry observes, to wish that the decisions of the Roman Catholic Church or its decision-making system ‘to be changed by secular authority is simply to want what Henry the Eighth achieved several centuries ago: the substitution of secular for religious authority. It is surely not to be wondered at if the courts decline to reproduce the work of the Reformation in the guise of enforcing a law against employment discrimination. Those who despair of the prospects of reform from within the Roman Catholic Church … always have the option of joining the Anglican Church.’ B Barry Culture and equality: An egalitarian critique of multiculturalism (2001) 176.

Act 4 of 2000. The reason the claim was brought under the Equality Act rather than the Employment Equity Act is presumably that the teacher concerned was considered an independent contractor rather than an employee. Sec 5(3) of the Equality Act provides that the Act ‘does not apply to any person to whom and to the extent to which the Employment Equity Act 1998 ... applies’. Sec 4(1) of the Employment Equity Act states that it applies to ‘employees’, where ‘employee’ is defined as excluding independent contractors. Had Strydom been considered an employee, presumably the Equality Act would have been inapplicable and the claim would have been brought under the Employment Equity Act.
to permit it to engage in work-related discrimination on the basis of sexual orientation with respect to a music teacher. The issue confronting the Court was whether a religious association's right to discriminate on prohibited grounds should extend beyond religious leaders to include a music teacher in a faith school not involved in religious instruction.

My purpose in this article is to determine the appropriate domain of anti-discrimination law in so far as government wishes to prevent work-related discrimination by, and within, religious associations. Claims by religious associations that they should be permitted to engage in work-related discrimination on otherwise illegal grounds give rise to a clash of rights: the rights to freedom of association and freedom of religion, on the one hand, and the right to equality on the other. My aim is to ascertain, with reference to Strydom, the circumstances in which it is acceptable to privilege the rights to religious liberty and freedom of association over the right to equality: circumstances, that is, in which religious associations should be permitted to engage in work-related discrimination on otherwise prohibited grounds.

2 Equality, freedom of association and religious liberty

The principle of equality, a core liberal right, requires that people should be treated with equal concern and respect. The inclusion of the right to equality in the Constitution obliges the state to ensure that the principle of equal treatment is not violated in central areas of people's lives such as work.

Work is central to most people because it provides an income stream, enabling individuals to house, feed and clothe themselves and their dependants. Work also provides non-economic benefits, such as a heightened sense of well-being, improved self-esteem and a reduced level of depression. The right to equality requires that all individuals have equal access to these benefits. Work opportunities should in general not depend on the characteristics of individuals, such as religion, race, gender

---

10 The right to freedom of association is protected in sec 18 of the Constitution, which provides that '[e]veryone has the right to freedom of association'. The right to freedom of religion is recorded in sec 15(1) of the Constitution, which reads: 'Everyone has the right to freedom of conscience, religion, thought, belief and opinion.' Freedom of religion and associational liberty are conjoined in sec 31(1): 'Persons belonging to a cultural, religious or linguistic community may not be denied the right, with other members of that community (a) to enjoy their culture, practise their religion and use their language; and (b) to form, join and maintain cultural, religious and linguistic associations and other organs of civil society.'

11 For a discussion of the right to equality in the context of South African constitutionalism, see C Albertyn & B Goldblatt ‘Equality’ in Woolman et al (n 1 above) ch 35.

Taking diversity seriously: Religious associations and work-related discrimination

Religious associations contribute to a system of economic opportunity. As a result, few exemptions from general public obligations are more significant than exceptions to anti-discrimination law in hiring, promotion and firing ... Religious associations are not only membership groups but also employers in the secular economy, whose practices affect non-members directly.

The state thus has a strong interest in eradicating unfair discrimination in their work-related practices.

What militates against the uniform enforcement of anti-discrimination legislation in respect of the discriminatory work-related practices of religious associations in all cases – and what limits the purview of otherwise applicable public principles generally – is that freedom of association and religious liberty are both core liberal values. Properly understood, liberalism does not insist that illiberal associations (including many religious organisations) must always ensure that the conduct of their internal affairs conforms to liberal principles, including the principle of non-discrimination, since this would be insufficiently solicitous of associational and religious liberty. Provided that membership of associations is voluntary and that freedom of exit is safeguarded, the state should as far as possible avoid requiring associations to achieve congruence between their activities and even core public principles such as non-discrimination. Rather, illiberal associations which are committed to violating the principle of equal respect – those whose norms mandate, for example, unequal treatment of men and women – should within limits be permitted the freedom to conduct their affairs in accordance with the wishes of their members. The workplace policies of religious associations may sometimes be unfairly discriminatory, yet the Constitution should be interpreted to say that in certain circumstances (which it is the aim of this article to specify) religious associations have the right to discriminate in their work policies on their own deplorable terms.

Why is freedom of association so important? One reason is that groups and associations contribute to their members' wellbeing. Freedom of association enables individuals to 'create and maintain intimate relationships of love and friendship' and is 'increasingly essential as a means of engaging in charity, commerce, industry, education, health care,

---

14 See Barry (n 7 above) 147 and W Galston ‘Two concepts of liberalism’ (1995) 105 Ethics 516 533.
15 Rosenblum (n 13 above) 167.
residential life [and] religious practice'.

Associational freedom is an essential part of individual freedom: associations represent the choices of their members about how to live. The value placed on freedom of association also reflects the importance of civil society, ‘that network of intimate, expressive, and associational institutions that stand between the individual and the state’, the discussion of which in the United States political tradition goes back to Alexis de Tocqueville. The existence of associations serves as a counterweight to potentially overweening state power.

In Roberts v United States Jaycees, the US Supreme Court recognised the right to freedom of association (which is not explicitly mentioned in the US Constitution, as it is in the South African Constitution) as limiting the extension to associations of otherwise applicable public principles. Brennan J expresses many of the strongest reasons for protecting freedom of association: It is ‘a fundamental element of human liberty’ that enables such activities as ‘speech, assembly … and the exercise of religion.’ Associations have ‘played a critical role in … cultivating and transmitting shared ideals and beliefs … and acting as critical buffers between the individual and the power of the state’ and individuals ‘draw much of their emotional enrichment from close ties with others’ forged in associations. Further, freedom of association ‘safeguards the ability independently to define one's identity that is central to any concept of liberty’.

Another way to capture the significance of freedom of association is to frame it as a matter of diversity. As Galston has argued, ‘properly understood, liberalism is about the protection of diversity’. A liberal state is one in which there exist associations embodying contrasting conceptions of the good life and differing views on what constitute worthy goals. As Brennan J puts it in Jaycees, associations ‘foster diversity’. He expands on this point as follows:

An individual’s freedom to speak [and] to worship … could not be vigorously protected from interference by the state unless a correlative freedom to engage in group effort toward those ends were not also guaranteed. According

---

17 G Kateb ‘The value of association’ in Gutmann (n 16 above) 36.
18 Galston (n 14 above) 531.
20 Roberts (n 19 above) 618.
21 Roberts (n 19 above) 618-619.
22 Galston (n 14 above) 523. To take a prominent example of modern liberal philosophy, John Rawls's Political liberalism (1993) is a response to what he calls 'the fact of reasonable pluralism' (36). It purports to take seriously 'the diversity of religious, philosophical and moral doctrines found in modern democratic societies' (136).
23 Roberts (n 19 above) 619.
24 Roberts (n 19 above) 622.
Taking diversity seriously: Religious associations and work-related discrimination

Associations should not always be expected to conform to public principles, including non-discrimination, when those principles clash with the convictions of members, and the state should refrain as far as possible from interfering with the internal affairs of associations. This is what the protection of diversity requires. Churches and church-run institutions and groups are not simply associations; they are religious associations. Claims by religious associations that they be permitted to discriminate in the workplace on otherwise prohibited grounds are typically grounded on the right to freedom of religion as well as the right to freedom of association: 'Religion includes important communal elements for most believers. They exercise their religion through religious organisations, and these organisations must be protected by the [Free Exercise] Clause.'25 The right to freedom of religion incorporates the interest of churches in conducting their internal affairs, including labour relations, according to their perceived interests; that is, in keeping decision-making authority over operations within the church free of government interference.26 The argument that government should not regulate the way in which churches select their officers and workers is that to do so is an unacceptable encroachment on religious affairs: It is simply inappropriate for the state to insist that religious organisations carry on their internal affairs in a way that is inconsistent with settled religious convictions and practices.27

The right to religious liberty, like the rights to equality and freedom of association, is a fundamental right. The plausibility of cognitive and normative claims made by religious believers need not be admitted to appreciate that the right to freedom of religion protects important interests. Religion is a significant part of believers' lives and their 'search for life's ultimate meaning'; it 'protects people's responsibility to find value in their lives.'28 It is a source of identity that is closely connected to self-respect and dignity, as well as a source of moral values. But 'one cannot conceive of meaningful individual religious freedom unless religious institutions are protected from government encroachment'.29 Indeed, once the interests protected by the right to freedom of religion are added to those protected by the right to associational liberty it may be that, as Greenawalt contends, claims by religious associations for exemption from anti-discrimination

26 Laycock (n 25 above) 1373 1414.
laws ‘will often have more force than claims deriving from other associations’.30

Neither freedom of religion nor associational liberty, however, are absolute rights. Nor is the right to equality an absolute right. Freedom of association and religious liberty will sometimes have to give way to the right to freedom from discrimination. Resolving in a particular case the question of whether the rights to associational freedom and religious liberty should prevail over non-discrimination depends on a balance: whether the state should be permitted ‘to subject religious institutions to laws that substantially burden those institutions, or even strike at their heart … depends on the extent of the interference with religious convictions and the strength of the state's justification’.31 It is in the nature of such a proportionality exercise that ‘neither conclusion (to interfere or not to interfere) can be justified by way of a presumption (for or against interference)’.32

3   Equality is not more fundamental than freedom of religion

In Strydom, Basson J creates the impression that the right to equality is more fundamental, important and ‘core’ than the rights to freedom of religion and freedom of association, with the result that the church's claim is weakened in advance of his carrying out of the balancing exercise. He does so in two ways. First, he mentions the right to freedom of association only once, in passing, in a paragraph in which he (correctly) dismisses Taylor v Kurtstag33 as unhelpful to the question of how to balance religious liberty and equality in Strydom – Taylor being concerned with the rights of a member of a church, rather than a non-member as in Strydom, and not being a case involving unfair discrimination.34 His failure to consider the right to freedom of association more fully and to explicate the content of this right in the present context has the effect of improperly diminishing its significance, tipping the scales in favour of the right to equality.

30 K Greenawalt ‘Freedom of association and religious association’ in Gutmann (n 16 above) 136.
32 Gutmann (n 16 above) 13.
33 2005 1 SA 362 (W).
34 In Taylor (n 33 above), an orthodox Jew sought to prevent the Beth Din, a Jewish ecclesiastical court established to resolve disputes between members of the orthodox Jewish community, from publishing a cherem, a notice excommunicating him from the faith. Following his divorce, the Beth Din had reached a determination directing him to pay maintenance in respect of his children. He had failed to abide by the order, subsequent to which the Beth Din had issued the cherem. The applicant objected to the publication of the cherem on the grounds that it constituted defamation and violated his right freely to associate with other members of his religious community. The court ruled that as an orthodox Jew, he was ‘required to comply fully with the dictates of Jewish law and society and the mandates of Jewish law and ethics’, which include the
Secondly, he seems to think that the right to equality is more fundamental than the right to freedom of religion in the post-apartheid constitutional order. He does not make this claim explicitly, but it is strongly implied by what he says: that although freedom of religion is an ‘important’ right, the right to equality is ‘foundational to our constitutional order’. He quotes *Minister of Education v Syfrets Trust Ltd* 36 to the effect that equality is ‘not merely a fundamental right; it is a core value of the Constitution’ that ‘goes to the bedrock of our constitutional architecture’.37 The claim that the right to equality takes precedence over religious liberty is also implied in a paragraph that appears to state, in the context of the balancing exercise, a presumption in favour of equality:38

The question remains whether the right to religious freedom outweighs the constitutional imperative that there must not be unfair discrimination on the basis of sexual orientation. The constitutional right to equality is foundational to the open and democratic society envisaged by the Constitution. As a general principle, therefore, the Constitution will counteract rather than reinforce unfair discrimination on the ground of sexual discrimination.

The trouble with the claim that the right to equality is more fundamental,

---

35 Strydom (n 8 above) paras 9 & 10.
36 2006 4 SA 25 (CC).
37 Strydom (n 8 above) para 10.
38 Strydom (n 8 above) para 14.
'core' or important than the right to freedom of religion is that it is false.\textsuperscript{39} The right to freedom of religion is 'one of the most basic rights human beings possess. It would not be a huge exaggeration to suggest that this freedom is the grain of sand in the oyster of politics around which the pearl of liberalism gradually formed.'\textsuperscript{40} Some would go further and claim that the right to freedom of religion takes precedence over the right to equality; that the right to religious liberty has 'logical priority' relative to other rights;\textsuperscript{41} that it is 'the ultimate freedom' and 'not merely one of many rights, but the prototypical human right'.\textsuperscript{42}

The correct position, however, is that expressed in the context of the Canadian Charter of Rights and Freedoms by Lamer CJ for the majority of the Canadian Supreme Court in \textit{Dagenais v Canadian Broadcasting Corp}:\textsuperscript{43}

A hierarchical approach to rights, which places some over others, must be avoided, both when interpreting the \textit{Charter} and when developing the common law. When the protected rights of two individuals come into conflict … \textit{Charter} principles require a balance to be achieved that fully respects the importance of both sets of rights.

Accordingly, there cannot be a presumption in favour of the right to freedom of religion or the right to equality antecedent to the balancing of these rights. It is illegitimate to assert a presumption either way 'without begging the question: which side carries the weight of argument in cases of conflict' between the values of free association and religious liberty and non-discrimination.\textsuperscript{44}

\textsuperscript{39} For one version of the claim that the right to equality is the most important constitutional value, see J Rutherford 'Equality as the primary constitutional value: The case for applying employment discrimination laws to religion' (1996) 81 \textit{Cornell Law Review} 1049. As I Benson 'The freedom of conscience and religion in Canada: Challenges and opportunities' (2007) 21 \textit{Emory International Law Review} 111 148 states, however, '[e]quality is not more important than freedom of religion'.\textsuperscript{40} W Galston \textit{The practice of liberal pluralism} (2005) 178.\textsuperscript{41} M McConnell 'Religion and constitutional rights: Why religious liberty is the “first freedom”' (2000) 21 \textit{Cardozo Law Review} 1243 1244.\textsuperscript{42} K Hasson 'Religious liberty and human dignity: A tale of two declarations' (2003) 27 \textit{Harvard Journal of Law and Public Policy} 81 88 89.\textsuperscript{43} [1994] 3 SCR 835 877.\textsuperscript{44} Gutmann (n 16 above) 11. A further argument may be open to those who seek to privilege equality over other rights. It may be claimed that the Equality Act itself privileges equality. On Woolman's reading of the Equality Act, '[t]he Act makes equality \textit{per se} an overriding important goal' and 'a trump' (Woolman (n 16 above) 34). Although it does not 'engage directly the complicated cluster of issues raised by conflicts between equality and association', the Equality Act 'poses an imminent threat to associational freedom' (Woolman (n 16 above) 36), including, presumably, to religious associations. To the extent that the Equality Act gives equality precedence over associational freedom and religious liberty, this is illegitimate.
4 What did Strydom hold?

In Strydom, the contract of Strydom, a music teacher working in an academy run by the Dutch Reformed Church – an ‘independent contractor’ and ‘not an employee of the church’ – was terminated by the church when it discovered that he was involved in a homosexual relationship. Strydom, who was not a member of the church, claimed that by terminating the contract for this reason, the church had unfairly discriminated against him on the grounds of sexual orientation, one of the grounds expressly prohibited by the Equality Act. The church averred in response (a contention that was not disputed) that it was a settled belief of the church that ‘persons of homosexual orientation must ... be celibate and cannot be involved in a homosexual relationship’ which, if engaged in, would ‘amount to a cardinal sin in view of the church’s teachings based on the Bible’. The church claimed that persons in positions of leadership are required to comport themselves in accordance with church doctrine, including its doctrines concerning homosexuality. His position, the church asserted, required him to conduct himself as a ‘role model’, which meant following an ‘exemplary Christian lifestyle’. His involvement in a homosexual relationship, the church claimed, rendered him unfit to teach since it was inconsistent with an exemplary Christian life. The church claimed that a teacher's involvement in a homosexual relationship would ‘set a bad example to his students’ and that a failure to terminate his contract could result in the church being seen to condone a homosexual relationship.

Basson J proceeded to decide the matter on the basis that ‘the right of equality must be balanced against the freedom of religion of the church’. It will be useful for the purposes of assessing the decision to identify and distinguish key steps in the Court's reasoning as it carries out this proportionality exercise. The following findings and determinations appear in the judgment: (1) that an exemption to permit a church-run institution to engage in work-related discrimination should only be granted if the activities performed by the individual against whom the church seeks to discriminate are sufficiently proximate to the beliefs of the religion; (2) that notwithstanding that Strydom worked as a teacher in a religious

45 Strydim (n 8 above) paras 1 & 20.
46 Strydim (n 8 above) para 12. Basson J does not mention the biblical sources on which the church relied, but biblical verses frequently interpreted as injunctions against active homosexuality include Leviticus 18:22, Leviticus 20:13, Romans 1:26-7, 1 Corinthians 6:9-10 and 1 Timothy 1:9-10. There is no indication from the judgment that the existence of a coherent theological basis for the school's stance on sexual morality was disputed, but a resolution of this dispute, had it arisen, would anyway have taken the court beyond its legitimate role.
47 Strydim (n 8 above) para 15.
48 Strydim (n 8 above) 16.
49 Strydim (n 8 above) paras 21 & 23.
50 Strydim (n 8 above) para 8.
academy, his position as a music teacher was not one of a religious or spiritual leader, and since he was not required to teach religious doctrine, his work was insufficiently close to the doctrinal core of the church to justify an exemption to the Equality Act to permit the church to engage in otherwise prohibited discrimination; (3) that Strydom’s position as a music teacher was analogous to ‘ordinary’ workers such as ‘a person in a homosexual relationship … employed or contracted to do typing work as a secretary’, and whose jobs, unlike those of spiritual and religious leaders, are excessively remote from the religious group’s religious beliefs and practices; (4) that, despite Strydom being a teacher in a church-run academy, since it was not part of the job description in his contract that he was required to lead an exemplary Christian life, he could not be sanctioned for failing to ‘set a good example’ to students in the sense of leading a lifestyle consistent with the beliefs and teachings of the church; (5) that Strydom ‘was not in a position of leadership’ relative to the students; (6) that the church had no reason to be concerned about Strydom having a bad influence on pupils since there was ‘not a shred of evidence that [he] wanted to influence the students or any other church member’ and that, on the contrary, ‘he wanted to keep his homosexual relationship to himself as he regarded it as a private matter’; (7) that the church’s anxiety about condoning (and being seen to condone) a homosexual relationship by failing to terminate Strydom’s contract was misplaced, since the church ‘could have stated that it was required by the Constitution that they [sic] not discriminate on the basis of a person’s sexual orientation when concluding a contract of work’; (8) that the impact of the discrimination on Strydom was ‘enormous’, since he had suffered financially and psychologically as a result of his contract being terminated; (9) that the burden on the church’s religious liberty of permitting Strydom to continue to provide teaching services under the contract was ‘minimal’; and (10) that in the circumstances the right to be free from unfair discrimination should prevail over the right to freedom of religion, and that accordingly the church was not entitled to an exemption from the Equality Act.

In what follows, I shall argue that steps (2), (3), (5), (6), (7) and (9) of the Court’s reasoning are either questionable or false. I shall defend (1), (4), (8) and the Court’s ultimate finding in favour of Strydom expressed in (10). I shall be concerned in particular to show that unqualified endorsement of

51 Strydom (n 8 above) para 24.
52 Strydom (n 8 above) para 22.
53 Strydom (n 8 above) para 32.
54 Strydom (n 8 above) para 22.
55 Strydom (n 8 above) para 24.
56 Strydom (n 8 above) para 25.
57 As above.
58 The Court awarded Strydom damages ‘for the impairment of [his] dignity and emotional and psychological suffering’ and an additional amount for loss of earnings, as well as ordering the church to respond unconditionally to Strydom (paras 37, 38 & 39).
the Court's reasoning is misplaced. In response to the dismay expressed by religious groups in response to *Strydom* – the Apostolic Faith Mission and the African Christian Democratic Party urging that ‘the right to equality should not trump the right to religious freedom’, and the moderator of the Dutch Reformed Church, Professor Piet Strauss, expressing the view that ‘a church should have the freedom to stipulate a lifestyle based on the Bible for its employees and members’59 – constitutional law expert Professor Pierre de Vos appears to have endorsed unqualifiedly all of the steps set out in (1) to (10) above in his blog *Constitutionally Speaking*, commenting as follows:60

A person employed as a typist, a gardener (or a music teacher for that matter) does not give religious instruction and forcing a church to employ such a person in no way forces the church to espouse views or practices that its religion frowns upon or even abhor [sic]. The religious freedom of the church is therefore not affected while the rights of all citizens remain protected. This seems like a sensible compromise also adopted by the Pretoria High Court. To have held otherwise would have … created a situation in which religious institutions are in effect above the law. So this judgment manages to balance the interest of the state against the interest of religious groups in a sensible and pragmatic fashion, refusing to sanction religious tyranny. In a country where we are building a culture of respect for diversity and difference churches can therefore not get a free pass. This is, after all, not Iran or Germany. We respect different belief systems and do not give preference to the bigoted and homophobic views held by some.

I believe (as De Vos does not) that the church's claim for an exemption from the Equality Act has, from the perspective of liberal constitutionalism, more to recommend it than is apparent from the *Strydom* decision, although I agree (as does De Vos) with the Court's ultimate finding in favour of *Strydom*. What principally distinguishes my position from that of Basson J and De Vos is that, despite the value they profess to place on diversity, I believe that the reasoning employed in *Strydom* in fact fails to show sufficient respect for diversity.

As well as enquiring whether the reasons Basson J provides for his conclusion are plausible, I shall contrast the Equality Court's reasoning in *Strydom* with the reasons provided by courts in the United States, Canada and the United Kingdom in cases in which associations have claimed a right to discriminate on prohibited grounds. With the exception of its engagement with a single Canadian decision, the Equality Court preterms a survey of the approaches of the US, UK and Canadian courts in comparable cases. Yet I believe that a study of the approaches adopted


in other jurisdictions can help us to discover arguments for or against exemptions to laws that prohibit work-related discrimination that might otherwise be overlooked but are worthy of consideration. What an analysis of the US and Canadian cases in particular shows, I think, is that certain of the views expressed by the Court in *Strydom* are mistaken and in need of revision.

My analysis of the decisions in these other jurisdictions will not be confined to cases that deal with work-related discrimination by religious associations. In order to discover the fullest range of reasons and views potentially applicable to the issue of whether religious associations should be permitted to engage in otherwise illegal work-related discrimination, my analysis will be extended to include cases in which religious and secular associations have claimed the right to discriminate in the workplace and other contexts.

5 Comparative perspectives

5.1 The United States

Though it may at first appear otherwise, the US Supreme Court has in a series of decisions held that there are limits to the government's ability to enforce even core public commitments such as non-discrimination when these principles collide with religious beliefs of organisations. In *Bob Jones University v United States*, the Supreme Court had to determine whether the Internal Revenue Service (IRS) could rescind the University's tax exempt status on the grounds that it prohibited inter-racial dating and marriage between students. Bob Jones University, though not affiliated with any religious denomination, taught and propagated fundamentalist Christian beliefs. The sponsors of the University believed that the Bible forbids miscegenation and instituted a disciplinary rule to give effect to this conviction. As a result of its racially-discriminatory policy, the IRS revoked the University's tax-exempt status. The Supreme Court held that the IRS was entitled to do so. Burger CJ ruled that the justification for tax exemptions was that 'the exempt entity confers a public benefit' but, since the University's racially-discriminatory policy 'violates a deeply and widely-accepted view of elementary justice', it 'cannot be viewed as conferring a public benefit'. Since a tax exemption is a privilege rather than a right, the IRS was entitled to withdraw the exemption. In response to the University's argument that the IRS's revocation of the tax exemption

---


63 *Bob Jones University* (n 62 above) 591 592 595.
Taking diversity seriously: Religious associations and work-related discrimination

violated its free exercise rights, Burger CJ held that ‘[t]he government has a fundamental, overriding interest in eradicating racial discrimination in education … [which] substantially outweighs whatever burden denial of tax benefits places on petitioner's exercise of their religious beliefs’. 64

Both the Supreme Court's reasoning and its conclusion seem correct. But if that is so, does it not support the finding in Strydom that the state's interest in eradicating sexual orientation discrimination outweighs whatever burden is placed on the church's religious freedom by prohibiting it from discriminating on prohibited grounds? It does not. The issue in Strydom was not the revocation of a state-conferred privilege, but whether a church should be sanctioned for failing to make its internal affairs conform to a (meritorious) public principle: non-discrimination. In Bob Jones University, the Supreme Court does not rule that a private, religious university may be forced to abandon its internal, racially-discriminatory policy on pain of civil or criminal sanction. Such a step, as Galston observes, 65

would have implicated basic rights of freedom of association and free exercise of religion that the court has long defined and protected, and it would have called for an entirely different legal analysis. Even employing the line of reasoning it used to settle Bob Jones, the court would almost certainly have found that unlike the mere revocation of tax exemption, a direct order to change the discriminatory policy would have failed the balancing test by imposing an 'undue burden' on petitioners' free exercise.

So, the approach followed in Bob Jones University is on closer inspection hardly supportive of the approach taken in Strydom. Extending the precedent of Bob Jones University to the facts of Strydom would mean not invalidating the church's discrimination on a prohibited ground, but burdening the church by removing all forms of otherwise applicable public encouragement and favour, including financial subvention and tax-exempt status. 66

Let us return to the claim implicitly made by the court in Strydom reflected in (1) above: that an exemption to permit a church-run institution to engage in work-related discrimination should only be granted if the activities performed by the individual against whom the church seeks to discriminate are sufficiently proximate to the beliefs of the religion. A US Supreme Court decision subsequent to Bob Jones University – Corporation of the Presiding Bishop v Amos 67 – is worth considering because it contradicts this claim and also because, unlike Bob Jones University, it deals with work-related discrimination.

64 Bob Jones University (n 62 above) 604.
65 Galston (n 40 above) 183.
66 See, for a version of this proposal, I Shapiro Democracy's place (1996) 244.
In the United States, provisions outlawing employment discrimination on the grounds of race, colour, religion, sex and national origin are contained in Title VII of the Civil Rights Act, 1964, the basic federal law against employment discrimination. Unlike the Equality Act, Title VII incorporates a statutory exemption (section 702) which initially permitted discrimination on the grounds of religion by religious associations in respect of religious activities only, but was extended in 1972 to cover all non-profit-making activities of religious associations. In Amos, a janitor employed for 16 years at a Mormon-run gymnasium was dismissed for his failure to meet the requirements for a certificate of eligibility for attendance at Mormon temples. The dismissal clearly constituted job discrimination on the basis of religious affiliation, prohibited under Title VII. The janitor challenged the constitutionality of the section 702 exemption, in as much as it permitted religious associations to discriminate on religious grounds when filling non-religious positions, a category into which his own job fell.

The District Court ruled that the amendment to the Title VII exemption was unconstitutional because it enhanced the ability of religious associations to further their beliefs in situations where ‘people's opportunity to earn a living is at stake’. It held that, since there was no clear connection between the business of the gym and the beliefs of the Mormon Church, and since none of the janitor's duties were ‘even tangentially related to any conceivable religious belief or ritual of the Mormon Church or church administration’, the section 702 exemption would be impermissibly broad if interpreted as covering the requirement that a janitor in a gymnasium must be a member in good standing of the Mormon church. The District Court's ruling is consistent with Basson J's holding in Strydom that certain jobs (typists, for example) are so distant from the doctrinal core of the church's activities as to render an exemption from anti-discrimination legislation unwarranted.

A unanimous US Supreme Court, however, declined to follow the District Court. In his concurring judgment, Brennan J noted that the section 702 exemption ‘necessarily has the effect of burdening the religious liberty of prospective and current employees’ since it coerces such individuals into ‘the choice of conforming to certain religious tenets or losing a job opportunity’. At the same time, he observed, ‘religious organisations have an interest in autonomy in ordering their religious affairs’. Resolving the clash between the rights of religious organisations and those of employees would require ‘ideally’ that ‘religious organisations should be able to discriminate on the basis of religion only

---

68 It is worth emphasising that this exemption only covers religious discrimination and not discrimination based on race, gender or sexual orientation.
70 Amos (n 69 above) 802.
71 n 67 above, 340.
72 As above.
with respect to religious activities, so that a determination should be made in each case whether an activity is religious or secular’, since ‘the infringement on religious liberty that results from conditioning performance of secular activity upon religious belief cannot be defended as necessary for the community's self-definition’.73

Brennan J nevertheless decided that such a case-by-case determination would be inappropriate since it would involve ‘ongoing government entanglement in religious affairs’, which could result in a religious organisation being ‘chilled in its free exercise activity’.74 He expands on this point as follows:75

While a church may regard the conduct of certain functions as integral to its mission, a court may disagree. A religious organisation therefore would have an incentive to characterise as religious only those activities about which there likely would be no dispute, even if it genuinely believed that religious commitment was important in performing other tasks as well. As a result, the community's process of self-definition would be shaped in part by the prospects of litigation.

Brennan J concluded that the ‘categorical exemption’ contained in section 702 (as amended) covering all types of church employees, no matter what the job, appropriately balances the right of employees to religious liberty and the autonomy of religious organisations.76 In contradistinction to the District Court, the Supreme Court upheld the constitutionality of section 702.

On the kind of reasoning followed by the Supreme Court in Amos, (1) above is false and the Equality Court in Strydom is mistaken to require that the church demonstrate that the relevant position is connected to its religious beliefs or practices, since this would place pressure on religious associations to predict which of its activities judges would consider regulable and which exempt – fear of potential liability would ‘chill’ associational life. On the Supreme Court’s approach, Basson J should have shown ‘deference’77 to the church’s view of what it considers to be a religiously related activity.

The approach of the Supreme Court is, however, incorrect. There is no good reason to think that if an exemption were restricted to religious activities, courts would restrict the group’s self-definition by imposing a narrow view of the religious.78 True, there is a always a danger that when deciding whether to grant or withhold an exemption, courts will employ

73 n 67 above, 343.
74 As above.
75 n 67 above, 343 - 344. White J, who delivered the opinion of the court, expresses the same view at 336.
76 n 67 above, 345 - 346.
77 n 67 above, 345.
78 Rosenblum (n 13 above) 176.
Chapter 3

criteria that are not part of the church's self-understanding and will fail to
give due consideration to the religious association's own account of the
religious foundation of an employee's activities. Even so, this does not
mean that religious associations should not be required to demonstrate to
the court that the position related to the unfair discrimination is sufficiently
closely connected to the doctrinal core of the church.

The Amos court mistakenly sought to avoid the task of distinguishing
between work that is religious in nature, and work that is not. It 'refused to
require the Mormon Church to demonstrate that discrimination in hiring
at its gym was a matter of faith or affected religious life in terms of doctrine,
ritual or authority'.79 By contrast, Basson J rightly held that religious
groups should not be permitted to engage in work-related discrimination
unless the job in question is sufficiently closely connected to the doctrinal
core of the religion. Permitting religious organisations to coerce an
economically-dependent employee or contract worker into making this
choice by insisting on compliance with its religious discipline is certainly
to impose a burden on him, and can only be justified if the job in question
is religiously based. This is so because (as Brennan J recognised) only
religious activity should be viewed as necessary for the community's self-
definition. Basson J's approach is in any case consistent with the approach
of courts in the US in cases involving employment discrimination on
grounds other than religion (the only group covered by the section 702
exemption) including race and gender: in cases in which religious
associations discriminate on one of these other grounds, they are regularly
called upon to demonstrate the religious nature of the activity.

A final decision of the US Supreme Court, Boy Scouts v Dale,80 shows
how far the court is prepared to go to protect the rights of associations to
discriminate on otherwise prohibited grounds. Dale, admittedly, is not, in
several respects, analogous to Strydom: It is not concerned with work-
related discrimination, but rather with discrimination against a member,
and it is concerned with discrimination by a non-religious, rather than
religious, association. It nevertheless has in common with Strydom that it
involves discrimination by an association on the grounds of sexual
orientation, a legally-prohibited ground. The Boy Scouts of America
(BSA), 'a private, not-for-profit organisation engaged in instilling its
system of values in young people',81 revoked the adult membership of an
assistant scout master when it learned that he was an avowed homosexual
and gay rights activist. The BSA claimed that the homosexual conduct of
Dale was inconsistent with the values it seeks to instil in young scouts –
much as the church in Strydom averred that Strydom's homosexual activity
was inconsistent with the values it was seeking to impart to pupils. Dale
was an adult scout leader whose role in the BSA included 'instructing

79 Rosenblum (n 13 above) 190.
81 Dale (n 80 above) 643.
Taking diversity seriously: Religious associations and work-related discrimination

[young members] in activities such as camping, archery and fishing', thereby 'inculcat[ing] them with the Boy Scouts' values – both expressly and by example'.

A narrow majority of the Court held that the presence of an avowedly gay assistant scout master would 'significantly burden the organisation's right to oppose or disfavour homosexual conduct' since his 'presence in the Boy Scouts would, at the very least, force the organisation to send a message, both to the youth members and the world, that the Boy Scouts accepts homosexual conduct as a legitimate form of behaviour'. Forcing the BSA to accept homosexual members would therefore 'interfere with the Boy Scouts' choice not to propound a point of view contrary to its beliefs'. The majority held that this was true even if the BSA encouraged scout leaders to 'avoid questions of sexuality and teach only by example'.

The majority held that the BSA's associational right to discriminate against homosexuals overrode New Jersey's interest, even if compelling, in eliminating such discrimination.

Although Dale was not engaged in remunerative work for the BSA, he was, like Strydom, an instructor of young people. The Supreme Court's ruling in Dale seems to contradict the Equality Court's reasoning as expressed in (6), (7) and (9) above. It says that the presence of an avowed homosexual in a position of trust and authority in the BSA, even if he did not advocate homosexual conduct in the organisation or seek to influence those over whom he exercised authority, would send a message to his charges and to the world that it regarded homosexual conduct as a morally-legitimate form of behaviour. The Equality Court explicitly denied as much. The majority in Dale also determined that to force the BSA to retain Dale as a member would 'significantly burden' its 'right to oppose or disfavour homosexual conduct'. The Equality Court denied this too.

The outlook of the minority in Dale is much closer to the Equality Court's view in Strydom, in as much as it held that forcing the BSA to retain Dale as a member does not impose a serious burden on the BSA, 'nor does it wish to force the BSA to communicate any message that it does not wish to endorse'. As in Strydom, the minority held that there was no evidence to suggest that Dale had sought to influence others concerning the morality of active homosexuality or that he had any intention of doing so. Equally,

---

82 Dale (n 80 above) 649.
83 Dale (n 80 above) 659.
84 Dale (n 80 above) 653.
85 Dale (n 80 above) 641.
86 Dale (n 80 above) 655.
87 Dale (n 80 above) 659.
88 Dale (n 80 above) 665. An issue which divided the majority and minority was whether the Boy Scouts had ever expressed a moral stance against homosexuality. In contrast with the majority decision of Rehnquist CJ, who found that the BSA had indeed expressed such a view, the dissenting opinion of Stevens J took the view that 'BSA
it was 'not likely' that the presence of a gay assistant scout master in the BSA 'would be understood to send any message, either to Scouts or the world'. On the other hand, discrimination on the grounds of sexual orientation has 'caused serious and tangible harm'.

5.2 Canada

The Equality Court in Strydom is right not only to insist that the work-related discrimination should be justified on the basis of tenets of faith, but also to distinguish between religious and non-religious activities, so that only activities which are closely related to the religious convictions and practices of the religious association are eligible for an exemption to anti-discrimination legislation. The Court is correct that a typist's activities, like those of a janitor in a gym, bear no significant relation to the group's religious beliefs and practices. But, although it was dealing with the revocation of the membership of an individual rather than with work-related discrimination against a non-member (as in Strydom), the Supreme Court's decision in Dale invites the question: Could not the position of a teacher, even one not carrying out religious instruction, in a religious school have arguable religious significance, since the pedagogical function goes to the heart of a religious school's operations and ethos? Decisions of the Canadian courts and boards of enquiry bear on this question.

In Canada, protection against unfair discrimination is provided by equality provisions contained in the legislation of the various provinces, with all provinces subject to the Canadian Charter of Rights and Freedoms, which includes a right to equality. Although equality provisions vary between provinces, most include the principle of non-discrimination, but provide some form of occupational exemption, typically where there is a bona fide occupational requirement. Canadian courts and boards of enquiry have considered cases in which religious occupational requirements may lead to discrimination on protected grounds, and have

never took any clear and unequivocal position on homosexuality' (676) and that its 'public posture' was 'welcoming of all classes of boys and young men' (672). In Strydom, by contrast, the claim of the church that it regarded homosexual activity as a 'cardinal sin' was uncontested. Stevens J asserted that '[t]o prevail in asserting a right of expressive association as a defence to a charge of violating an anti-discrimination law, the organisation must at least show that it has adopted and advocated an unequivocal position inconsistent with a position advocated or epitomised by the person whom the organisation seeks to exclude' (687). The Equality Court found that the church had demonstrated at least this much.

Dale (n 80 above) 697.

Dale (n 80 above) 700.

See, for a decision reflecting this view, EEOC v Pacific Press Pub Assoc 676 F 2d 1272 (9th Cir, 1992). (The US Court of Appeals for the Ninth Circuit held a non-profit publishing house associated with the Seventh-Day Adventist Church liable for gender discrimination on the grounds that the complainant's duties as a typist were remote from the core of the operations of the religious organisation.) For a discussion of this and related decisions, see MA Hamilton God vs the gavel: Religion and the rule of law (2005) 173 - 199.
permitted religious associations to dismiss employees who have not conformed to the religious lifestyle stipulated by the employer.

In *Caldwell v St Thomas Aquinas High School*, the Canadian Supreme Court found for the school on both counts. First, it contended that section 8 of the British Columbia Human Rights Code (the anti-discrimination provision) should not operate to protect Mrs Caldwell if it was found that the non-renewal of her teaching contract resulted from her lacking a *bona fide* qualification. The school argued that Mrs Caldwell lacked a *bona fide* requirement for the job, as she was unwilling to observe the requirements and practices of the church. This requirement, it claimed, was necessary for upholding the religious nature of school, where teachers led pupils by example. Secondly, the school claimed that the discrimination was covered by section 22 of the Code, which exempts non-profit religious organisations from the Code by permitting them to prefer in their employment practices members of their religious group.

The Canadian Supreme Court found for the school on both counts. It observed that the case implicated conflicting rights, ‘that of an individual to freedom from discrimination in employment, and that of a religious group to carry on its activities in the operation of its denominational school according to its religious beliefs and practices’. The Court held that the dismissal of Mrs Caldwell was not a contravention of the Code; the school was entitled to impose the faith requirement and could discriminate against her on the basis of her failure to comply with the tenets of the church. It found that ‘[t]he relationship of the teacher to the student enables the teacher to form the mind and attitudes of the student’ and that there was nothing objectionable in the school’s requiring teachers ‘to observe and comply with the religious standards and to be examples in the manner of their behaviour in the school so that students see in practice the application of the principles of the church on a daily basis and thereby receive what is called a Catholic education’.

Addressing the contention that the requirement of religious conformity was not reasonably necessary to ensure the efficient

---

92 [1984] 2 SCR 603.  
93 *Caldwell* (n 92 above) 606.  
94 *Caldwell* (n 92 above) 618.
performance of the teaching function, the Court found that\textsuperscript{95}

the religious or doctrinal aspect of the school lies at its very heart and
colours all its activities and programs. The role of the teacher in this respect is
fundamental to the whole effort of the school, as much in its spiritual nature
as in the academic.

In these circumstances, the requirement of religious conformity was held
to be a \textit{bona fide} qualification.

The Court ruled further that the dismissal of Mrs Caldwell was a
permitted preference under section 22 of the Code, holding that this
section confers on faith schools\textsuperscript{96}

the right to preserve the religious basis of the schools and in so doing engage
teachers who by religion and by the acceptance of the church's rules are
competent to teach within the requirements of the school. This involves and
justifies a policy of preferring Roman Catholic teachers who accept and prac-
tice the teachings of the church.

The approach followed in \textit{Caldwell} would seem to contradict (2), (6) and
(9) of the Equality Court's reasoning in \textit{Strydom}. Basson J contrives to
distinguish \textit{Caldwell} for three reasons, but two of these are unconvincing.
The first ground on which he says that \textit{Caldwell} can be distinguished is that
'the teaching of doctrine and the observance of standards by teachers formed part of the contract of
employment of teachers'.\textsuperscript{98} Although this is
indeed so, this fact does not have the effect of rendering several of the views
expressed in \textit{Caldwell} inapplicable to Strydom's case. The issue facing the
Equality Court in \textit{Strydom} is whether the right to freedom of religion may
be invoked to justify otherwise forbidden discrimination against a teacher
in a faith school. On the view expressed by McIntyre J in \textit{Caldwell}, such
discrimination is justified. Several of the reasons he provides in support of
this view have application in the circumstances of \textit{Strydom}.

Secondly, states Basson J, in \textit{Caldwell} 'the teaching of doctrine and the
observance of standards by teachers formed part of the contract of
employment of teachers',\textsuperscript{98} whereas Strydom was not involved in religious
instruction, which was carried out by religious leaders in the church.\textsuperscript{99}
Although Basson J is correct that at the school at which Mrs Caldwell
taught, the teaching of doctrine was included in teachers' employment
contracts, the teaching of religious doctrine is not an issue that divides the
two cases since (as Basson J neglects to acknowledge) Mrs Caldwell taught
only commercial subjects and mathematics and, like Strydom, 'did not

\begin{itemize}
\item \textsuperscript{95} \textit{Caldwell} (n 92 above) 624.
\item \textsuperscript{96} \textit{Caldwell} (n 92 above) 628.
\item \textsuperscript{97} \textit{Strydom} (n 8 above) para 27.
\item \textsuperscript{98} As above.
\item \textsuperscript{99} \textit{Strydom} (n 8 above) para 28.
\end{itemize}
give any formal religious instruction’ apart from leading prayers, as did other teachers, including non-Catholics.100

A third distinction between the two cases asserted by Basson J was that in the school at which Mrs Caldwell taught it was part of teachers’ job descriptions that each was ‘expected to be an example consistent with the teachings of the Church, and must proclaim Catholic philosophy by his or her conduct within and without the school’.101 By contrast, Basson J found that the church had failed to show, in the absence of a written contract, that it was part of Strydom’s job description to become a role model for Christian values. This is a genuine point of divergence between the two cases.

A similar decision to Caldwell was reached in a subsequent case involving dismissal for failure to comply with the standards set by a religious school. In Garrod v Rhema Christian School,102 a teacher in a Christian school was dismissed for having an extra-marital relationship, which the school claimed ‘was in flagrant conflict with the principles of Christian morals’.103 The teacher claimed that she had been discriminated against on the grounds of marital status. The Ontario Board of Enquiry determined that in these circumstances the religious right of the employer to require its staff to lead exemplary Christian lives trumped the rights of the employee not to be discriminated against on grounds of marital status. The teacher concerned acknowledged that it had been made clear to her that being a teacher at the school required her to be a role model for students.104 There is no evidence in the decision that the teacher had participated in the religious instruction of pupils.

In another decision, Kearley v Pentecostal Assemblies Board of Education,105 a divorced teacher at a Pentecostal collegiate resigned under protest when the collegiate indicated it intended to fire her. She had remarried in contravention of the church’s belief that divorced persons should not remarry while their former spouses are still living. She claimed that the school had discriminated against her on the basis of marital status and religion. She had contracted to abide by guidelines for Christian teachers, which state that teachers are to support and uphold the basic doctrines of the Collegiate106 and ‘to demonstrate, through precept and

100 Caldwell (n 92 above) 606.
101 Caldwell (n 92 above) 608.
103 Garrod (n 102 above) para 29.
104 Garrod (n 102 above) para 14.
106 Kearley (n 105 above) para 31.
example, the virtues of a Christian life’. Following *Caldwell*, the Newfoundland Board of Enquiry concluded that requiring teachers’ conformity to the religious beliefs and rules of the church was reasonably necessary to protect the religious nature of the school. Again, there was no evidence that the teacher concerned was involved in religious instruction.

Canadian decision makers have not always found in favour of religious associations that engage in employment discrimination on otherwise prohibited grounds, however. In *Parks v Christian Horizons*, a carer in a group home run by a Christian organisation claimed she had been discriminated against on the grounds of marital status after she was dismissed for being involved in an extra-marital relationship. By contrast with *Caldwell, Garrod* and *Kearley*, ‘there was no mention at any time of the requirement that all employees had to adopt a moral and sexual lifestyle that would not be in conflict with the Evangelical Christian doctrinal principles of Christian Horizons’. The employment contract did not include a provision imposing this requirement. The Court found for the employee, reasoning that employment discrimination of this kind would only be acceptable if ‘lifestyle requirements [were] clearly indicated or referred to in the employment contracts, and if at all possible, confirmed in the application and interview process leading to employment’.

Following the decision in *Parks*, Christian Horizons introduced a ‘Lifestyle and Morality Statement’ for inclusion in all contracts of employment. Nevertheless, in its recent decision, *Heintz v Christian Horizons*, the Human Rights Tribunal of Ontario refused to permit Christian Horizons to discriminate against an employee, Heintz, by terminating her employment when it discovered she was a lesbian involved in a same-sex relationship. Christian Horizons justified its discrimination on the grounds that Heintz’s conduct was inconsistent with the fundamental Evangelical Christian principles recorded in the ‘Lifestyle and Morality Statement’ and incorporated into her employment contract. The Tribunal found against Christian Horizons on the ground that it had not demonstrated that the discrimination was reasonably necessary.

Notwithstanding the Tribunal’s finding in favour of Heintz, the *Heintz* decision does not alter the Canadian jurisprudence dealing with religious educational institutions since the Tribunal affirmed the decisions in *Caldwell* and *Garrod*. It distinguished these decisions from *Heintz* on the grounds that, unlike the schools in *Caldwell* and *Garrod*, Christian Horizons was not in the business of religious education, and did not provide services

107 *Kearley* (n 105 above) para 54.
108 *Parks* (n 108 above) para 7.
109 *Parks* (n 108 above) para 8.
110 *Parks* (n 108 above) para 57.
to co-religionists exclusively. So Heintz is not inconsistent with the rule established in Caldwell, Garrod and Kearley that a religious school may reasonably require its teachers to abide by a lifestyle requirement that discriminates on otherwise prohibited grounds and by dismissing teachers who fail to do so.

5.3 The United Kingdom

In the UK, the Human Rights Act, 1998 incorporates a right to religious liberty and a right to be free from discrimination (articles 9 and 14 respectively of the European Convention on Human Rights (European Convention)). Exemptions to legislation that outlaws discrimination on the grounds of sex, religion and sexual orientation permit religious employers to discriminate on these grounds in certain instances.

Discrimination in employment on the grounds of religion is prohibited by the Employment Equality (Religion or Belief) Regulations, 2003, which includes an exemption for employers who have ‘an ethos based on religion or belief’. Regulation 7(3) provides that such employers can discriminate on the grounds of religion where being of a particular religion or belief is ‘a genuine occupational requirement for the job’ and ‘it is proportionate to apply that requirement in the particular case’.

Discrimination in employment on the grounds of sex is outlawed by the Sex Discrimination Act, 1975, as amended by the Employment Equality (Sex Discrimination) Regulations, 2005. Section 19 of this Act, as amended, provides that sex discrimination in cases in which ‘employment is for the purposes of an organised religion’ is permissible, provided that the requirement that employees be of a particular gender is imposed on one of two bases: where the discriminatory requirement is imposed ‘so as to comply with the doctrines of the religion’, or where the requirement is imposed ‘because of the nature of the employment and the context in which it is carried out, so as to avoid conflicting with the strongly held religious convictions of a significant number of the religion's followers’.

The Employment Equality (Sexual Orientation) Regulations, 2003 prohibit discrimination on the grounds of sexual orientation, but include an exemption (Regulation 7(3)) resembling the exemption for discrimination on the grounds of sex: ‘If the employment is for the purposes of an organised religion’ a requirement related to sexual orientation may be imposed either if it is needed ‘to comply with the
doctrines of religion’ or ‘because of the nature of the employment and the context in which it is carried out, so as to avoid conflicting with the strongly-held convictions of a significant number of the religion’s followers’.

The exemption to permit discrimination on the grounds of sexual orientation has proved problematic, however. In *R (Amicus MSF Section) v Secretary of State for Trade and Industry*,117 several provisions of the 2003 Employment Equality (Sexual Orientation) Regulations were challenged on the basis that they were too wide and so in breach of the European Convention. Upholding the Regulations, Richards J found on the basis of parliamentary statements that the Regulation 7(3) exemption was intended to be ‘very narrow’, since it is a derogation from the principle of equal treatment.118 He was of the view that the term ‘organised religion’ referred to in Regulation 7(3) was narrower than ‘religious organisation’: employment as a teacher in a faith school, he stated, is likely to be ‘for the purposes of a religious organisation’ but not ‘for the purposes of an organised religion’.119 In response to the averment that the Regulation 7(3) exemption authorises discrimination in cases including that of a ‘school for girls managed by a Catholic order [that] dismisses a science teacher on learning that she has been in a lesbian relationship, reasoning that such a relationship has been contrary to the doctrines of the Order’,120 Richards J held that on the narrow construction of the exemption favoured by him, the exemption would not cover this case.121 As Vickers comments, ‘[t]he legislation as interpreted in *Amicus* provides a clear steer that religious convictions can be used to justify discrimination on grounds of sexual orientation or gender, but only in the strictly limited circumstances of the appointment to clergy’.122

*Amicus* seems to support the decision of the Equality Court in *Strydom*, but there are reasons to question both the approach followed and its applicability to *Strydom’s* case. The reasoning of the UK High Court rests substantially on the wording of the section 7(3) exemption, which is expressed to be applicable to ‘organised religions’ but not to ‘religious organisations’. Richards J held the former to be narrower than the latter, such that although an ‘organised religion’ will also be a ‘religious organisation’, a ‘religious organisation’ will not always be an ‘organised religion’. But excluding religious organisations from the ambit of

---

118 *R* (n 117 above) para 115.
119 *R* (n 117 above) para 116.
120 *R* (n 117 above) para 95.
121 *R* (n 117 above) para 121.
consideration for possible exemptions is illegitimate. As Sandberg and Doe observe:

Given that Parliament in enacting section 13 of the Human Rights Act 1998 has made it clear that particular regard be given to the rights of a 'religious organisation', it is appropriate to ask why certain privileges have been afforded only to the narrower category [of 'organised religion']. This may actually infringe not only section 13 but more importantly article 9 of the ECHR.

The narrow construction of the exemption as expressed by Richards J fails to protect religious liberty and freedom of association adequately. It also seems to reflect an unwarranted presumption in favour of equality. Since freedom of religion cannot meaningfully be protected unless ‘religious organisations' are protected from government encroachment, and because a presumption in favour of equality and against freedom of religion is illegitimate, South African courts should decline to follow Amicus.

6  Was Strydom correctly decided?

The Equality Court's finding that Strydom suffered serious injury as a result of the church's discrimination – (8) above – cannot be gainsaid: Apart from the violation of his dignity inherent in being the victim of such discrimination, 'he suffered depression and was unemployed' and 'had to sell his piano and house'.

The Equality Court nevertheless underestimates the reasons for thinking that the burden placed on a church by the anti-discrimination legislation may be substantial. To disallow a church from discriminating impairs the ability of the religious community of which it forms a key part to transmit its core beliefs - including the belief that homosexual activity is sinful – by example, and may also impair the ability of a church to maintain the religious ethos of its academy, which includes the exemplification of these beliefs in its practices. This threatens religious pluralism and diverse civil associations, which work to express a range of conceptions of the good life and mitigate state power.

---

123 R Sandberg & N Doe 'Religious exemptions in discrimination law' (2007) 66 Cambridge Law Journal 302 310. Sec 13 of the Human Rights Act reads: 'If a court's determination of any question arising under this Act might affect the exercise by a religious organisation (itself or its members collectively) of the Convention right to freedom of thought, conscience and religion, it must have particular regard to the importance of that right.' The Human Rights Act incorporates into English law art 9 of the European Convention, which reads as follows: 'Everyone has the right to freedom of thought, conscience and religion; this right includes freedom to change his religion or belief, and freedom, either alone or in community with others and in public or private, to manifest his religion or belief, in worship, teaching, practice and observance.'

124 Strydom (n 8 above) para 33. On the relationship between equality and dignity, see Albertyn & Goldblatt (n 11 above) 8-13.

125 Galston (n 14 above) 533.
The reasons for thinking that the burden placed on the church may have been substantial need further discussion. The first reason to think so is the idea that churches transmit beliefs and moral values through the example set by teachers in church-run educational institutions. In *Strydom*, the church's position reflects the Aristotelian (or aretaic) view that moral virtue is not simply taught, but is acquired by pupils through their association with teachers who are themselves virtuous, with the corollary that it is wrong to place pupils with teachers who are not virtuous, because pupils might thereby be led to conclude that non-virtuous conduct (including active homosexuality, from the church's perspective) is morally acceptable.126 It is in the light of such a theory that the church's concern makes sense, that its teachers should be role models (in the sense of leading exemplary Christian lives) and that Strydom's homosexual lifestyle would set a bad example to students. In the Aristotelian conception, teachers teach moral values not didactically, as in the case of arithmetic, but through example. In this view, it makes no sense to say, as does Basson J, that the church's concern about Strydom's setting a bad example is misplaced because there was no evidence that he wanted to influence students.127 Irrespective of his intention not to influence students in matters of sexuality, in the Aristotelian view his example is itself influential. As Basson J acknowledges, it is not far-fetched to view Strydom as a 'mentor of the students on a personal ... level'.128 A mentor, according to the OED, is someone who 'acts as a guide and adviser to another person, especially one who is younger and less experienced'. Strydom's position might not have been one of a spiritual or religious leader or instructor, but as a mentor who acts as guide and adviser to those younger and less experienced, it is hard to accept Basson J's contention (reflected in (5) above) that, at least in relation to his pupils, he 'was not in a position of leadership'.129

Courts appear to have accepted that teachers exert an influence over pupils not only through the advice or instruction they give, but also in virtue of the example they set. As the Canadian Supreme Court observed in *Ross v New Brunswick School District No 15*,130 '[t]eachers occupy positions of trust and confidence, and exert considerable influence over their students as a result of their positions'.131 The Court quotes Canadian legal academic Allison Reyes:132

> Teachers are a significant part of the unofficial curriculum because of their status as ‘medium’. In a very significant way the transmission of prescribed

---

126 For a version of this view of moral virtue, see I Scheffler *Of human potential: An essay in the philosophy of education* (1985).
127 *Strydom* (n 8 above) para 22.
128 As above.
129 As above.
131 *Ross* (n 130 above) para 43 (my emphasis).
132 As above.
Taking diversity seriously: Religious associations and work-related discrimination

‘messages’ (values, beliefs, knowledge) depends on the fitness of the ‘medium’ (the teacher).

La Forest J adds:\(^{133}\)

By their conduct, teachers as ‘medium’ must be perceived to uphold the values, beliefs and knowledge sought to be transmitted by the school system. The conduct of a teacher is evaluated on the basis of his or her position, rather than whether the conduct occurs within the classroom or beyond … teachers do not necessarily check their teaching hats at the school yard gate and may be perceived to be wearing their teaching hats even off duty.

If it is true that teachers transmit beliefs and moral values through their example, then it would appear that Basson J’s contention in (6) above is incorrect, since even those teachers in a religious school who are not involved in religious instruction may exert an influence over children which is ‘bad’ from the perspective of the church.

Basson J may be mistaken to assert (3) above: that the position of a teacher involved in the teaching of a non-religious subject is analogous to those of employees like typists. The position of a teacher is not analogous to a typist in one crucial respect, which was expressed by the Ottawa Board of Enquiry in a matter concerning a complaint arising from the refusal of the Catholic School Board of Ottawa to engage a non-Catholic as a secretary for clerical duties in the school administration: In the Matter of the Ontario Human Rights Code and In the Matter of the Complaint of Mrs Bonnie Gore. McIntyre J in Caldwell quotes from the decision in that matter as follows:\(^{134}\)

I cannot see how a secretary can be expected to provide an example for the children. This is surely the responsibility of teachers, and the religious aspect is the responsibility of the ecclesiastics as well as most of the teachers. The secretary performs secretarial and clerical functions. Requiring that she be a Roman Catholic is not, in my opinion, a ‘reasonable occupational qualification’.

As McIntyre J notes, the Gore decision ‘recognised the special nature of the Catholic school, but concluded that the clerical worker would not have that degree of contact with the pupils that would make it essential in the interests of the church to have a Catholic in such position’.\(^{135}\)

Typists do not stand in the same relation of authority and mentorship relative to pupils as teachers do. The work of typists in a religious school is to type; the work of teachers in a religious school includes transmitting the beliefs and values of the school, didactically in the case of those involved

\(^{133}\) Ross (n 130 above) para 44 (my emphasis).

\(^{134}\) Caldwell (n 92 above) 619.

\(^{135}\) Caldwell (n 92 above) 620.
in religious instruction and by example in the case of all teachers. This is why it is correct to say that the work of typists or janitors is distant from the religious beliefs of the religious association for which they work, but that the activities of teachers of non-religious subjects in a religious school have a close connection to the religious beliefs of the church that runs the school. On this argument, an exemption in respect of all teachers, even those not involved in religious instruction, may be justified, but an exemption in respect of typists and janitors will not be.

Having set out the ‘role model’ argument, I confess to being unsure as to what weight to accord it, for the objection may be raised that it should not be accepted without empirical evidence. In the circumstances of Strydom, it may be argued that a court is entitled to expect compelling evidence to be presented before accepting that a gay teacher who has no desire to discuss, let alone defend, his homosexual lifestyle will influence pupils into thinking that his lifestyle is morally acceptable. As the Ontario Board of Enquiry observed in Garrod,\textsuperscript{136}

much depends on whether [the] role model theory actually has an effect on students, particularly given the external forces that inevitably impinge on their consciousnesses. No evidence of … studies was introduced, counsel preferring to rely mainly on logical and impressionistic argument. Nor was any evidence introduced of the force of the connection claimed between teachers’ personal and occupational lifestyles.

Let us consider a second reason for thinking that the church might be burdened significantly were it to be prevented from discriminating. The church expressed concern that to have kept Strydom on in a teaching position would have given the impression that it condoned homosexual relationships.\textsuperscript{137} If the church is correct that such an impression would be created, this is a serious concern – one accepted as important and valid by the majority of the Supreme Court in Dale. The appearance of condonation would surely undermine the teaching of a belief that active homosexuality is sinful, as well as sending out a misleading and potentially confusing message to the community as a whole.

It is an inadequate response to the church’s anxiety that, by retaining Strydom’s services after it came to light that he was in a homosexual relationship, it could be viewed as condoning homosexual activity to say, as Basson J does, that ‘they [sic] could have stated that it was required by the Constitution that they [sic] not discriminate on the basis of a person’s sexual orientation when concluding a contract of work’.\textsuperscript{138} Basson J’s suggestion is question-begging, for it assumes precisely what the church disputes and what is at issue in this case. It is the church’s claim that the

\textsuperscript{136} \textit{Garrod} (n 102 above) para 140.
\textsuperscript{137} \textit{Strydom} (n 8 above) para 23.
\textsuperscript{138} \textit{Strydom} (n 8 above) para 24.
Constitution, properly interpreted, requires that religious liberty and associational freedom should in the present case prevail over Strydom's right to be free from discrimination.

Is the church's concern valid that it might be viewed as condoning an actively homosexual lifestyle by retaining Strydom's services? I think there is reason to take it seriously, given that the church had placed Strydom in a position of authority and trust relative to students (as typists and janitors are not), on the basis of its belief that he was not committed to a life of cardinal sin. For the church to retain the services of an individual in such a position after it is discovered that he leads a homosexual lifestyle may well create an impression that it no longer holds the belief that active homosexuality is a mortal sin. Whether it does so may depend on whether Strydom's students and the members of the community of which the church forms a part were aware of his sexual orientation, a matter about which the Strydom judgment is silent.

There is a further related (and, I think, persuasive) reason to think that refusing to permit the church to discriminate against a teacher like Strydom would be burdensome, and disallowing discrimination against typists and janitors would not. To require that the church not discriminate against an actively homosexual teacher is to impair its ability to maintain the ethos of the academy: its religious character and identity. For the church to keep on an actively homosexual teacher would alter the defining character of the academy, its prevailing tone of sentiment, which includes a view that the homosexual lifestyle is sinful. Teachers, even those teaching only purely non-religious subjects, serve the academy's core purpose, which is pedagogical; they are defining and representative of the institution's religious ethos. Parents should be entitled to send their children to an educational institution with a particular religious ethos and such an institution should be permitted to require that teachers exemplify the tenets of the church in their conduct. By contrast, typists and janitors are not involved with the academy's core pedagogical activity and are not representative of the academy in the same way: an actively homosexual typist will not significantly compromise or alter the academy's character or identity in the way that an actively homosexual teacher might.

A determination that religious associations should be permitted to discriminate against teachers not tasked with religious education in order to safeguard the religious ethos of the school accords with the Canadian jurisprudence examined above: In the three decisions in which courts ruled that schools should be permitted to engage in employment discrimination against teachers (Caldwell, Garrod and Kearley), the religious ethos of the school was an important factor in the court's reasoning. In none of these decisions was there any indication that the teachers concerned taught religious instruction.
A church should, to protect the religious ethos of the school, be permitted to require as a condition of work that teachers, even those not involved in religious instruction, lead a religiously righteous lifestyle – one that excludes behaviour that violates the tenets of the church – even if this requirement effectively blocks certain work opportunities for homosexuals.\(^{139}\) Does this mean that the Equality Court's ultimate finding in favour of Strydom ((10) above) was mistaken? No, because although a school should be permitted to stipulate a lifestyle requirement for teachers, in view of the seriousness of work-related discrimination it should expressly be drawn to teachers' attention at the time of their hiring that it will be part of their job description to act as a role model for Christian values, and teachers should be required to assent to this requirement, preferably in the form of a provision to this effect in their work contracts. In Caldwell, Garrod and Kearley, the teachers concerned all acknowledged that this requirement had been made clear to them, while in Caldwell and Kearley the teachers contracted themselves to live in a Christian manner. In Parks, by contrast, the main reason the Court found in favour of the teacher was that the lifestyle requirement had not been clearly imposed by the church in advance, and did not appear in the employment contract.

Basson J found on the evidence that, although Strydom was ‘questioned on his Christian values in relation as [sic] to whether he has a personal relationship with God’ during his interview for the position of music teacher, it was not ‘part of his job description that he was to become a role model for Christianity’.\(^{140}\) The church might well have thought (it is not clear from the judgment) that it was an implied term of his contract that Strydom would conduct himself according to the tenets of the church. But this is insufficient: It should have been included as an express term. Assuming that the Court is correct that it was not made clear to Strydom that it was part of his job description to comport himself in accordance with the beliefs of the church ((4) above), its finding against the church ((10) above) is correct.

Had the facts of Strydom been slightly different, however – had the church expressly imposed a requirement on Strydom that he act as a role model for Christianity at the time that his contract was entered into and had it been explained what such a requirement entails – then an exemption permitting the church to discriminate on the grounds of sexual orientation would have been justified. There is no evidence to suggest that Strydom

\(^{139}\) For a homosexual, such a requirement makes working at such organisations conditional upon celibacy. Homosexuals could, of course, conform their behaviour to this stipulation, but the cost of doing so for most will be sufficiently high effectively to block these work opportunities. Not only is the repression of sexual impulses a denial of liberty; but it also, as HLA Hart *Law, liberty and morality* (1962) 21 understood, ‘creates a special and chronic form of suffering because of the nature of the desires it frustrates, impulses which affect “the development or balance of the individual’s emotional life, happiness and personality”’.

\(^{140}\) *Strydom* (n 8 above) paras 19 & 22.
was deprived of alternative work opportunities. His unemployment following the termination of his contract by the church was not the result of there being no other work opportunities available to him, but was instead due to the publicity that he ‘allowed his case to receive’. Had the church stipulated a Christian lifestyle requirement from the beginning then, assuming alternative work opportunities would have been available to him, the burden on Strydom would not have been ‘enormous’. As Vickers notes, if an individual ‘is unable to take up a job because she cannot meet its requirements, but remains free to take other jobs, there will only be a minor infringement of her interests, as she remains free to take up other work’.

Liberals may well find the church’s discriminatory policy reprehensible, as I do, but a liberal pluralist politics and jurisprudence should acknowledge the substantial burden placed on a faith school by preventing it from discriminating in accordance with the tenets of its faith.

Nevertheless, although committed to the view that such work-related discrimination should be legally permitted, I think it legitimate to oppose such discrimination by other means. What the Bob Jones decision of the US Supreme Court shows is that associations conducting their affairs in a manner contrary to core public purposes can legitimately be burdened. That is, the state is entitled to exert moral and financial pressure on a church to rethink its discriminatory policy by rescinding all forms of otherwise available public encouragement and favour, including a tax exempt status (if applicable) and all forms of financial subvention.

7 Conclusion

As McLachlin J put it in *Trinity Western University v College of Teachers*, ‘[t]he diversity of Canadian society is partly reflected in the multiple religious organisations that mark the societal landscape and this diversity of views should be respected’. In *Strydom*, the Equality Court likewise acknowledges the need to respect diversity in South Africa, quoting with approval Ngcobo J’s dissenting judgment in *Prince v President of the Law Society of the Cape of Good Hope*.

The right to religious freedom is especially important for our constitutional democracy. Our society is diverse. It is comprised of men and women of different cultural, social, religious and linguistic backgrounds. Our constitution recognises this diversity. The protection of diversity is the hallmark of a free and open society.

---

141 Vickers (n 122 above) 60.
143 *Prince* (n 1 above).
144 *Strydom* (n 8 above) para 9.
Yet the Equality Court’s reasoning in *Strydom* – including, principally, its privileging of equality over freedom of religion and associational liberty, and its trivialising of the burden placed on the church by continuing with Strydom’s services – reflects a failure to take diversity sufficiently seriously, even if its conclusion is correct.

Taking diversity seriously entails that the state does not have the right to abolish unfair discrimination in all its forms wherever it might appear: The right of a liberal state to enforce public principles such as non-discrimination is limited by the rights to freedom of association and freedom of religion. Latitude should be accorded to religious associations to allow them to govern their internal affairs and that includes accommodating their otherwise illegal work-related discrimination. At the same time, a religious association does not have the right to discriminate on otherwise prohibited grounds in respect of all activities performed by its employees and contract workers: The right to equality is after all a fundamental right, so that, as Richards J correctly observes in *R*, ‘the weight to be given to religious rights may depend upon how close the subject matter is to the core of the religion’s values or organisation’. Indeed, this is what it should depend on, for religious associations contribute to the system of economic opportunity that affects non-members directly.

Religious bodies should not be permitted to engage in work-related discrimination where the activity to be performed by the employee or contract worker or prospective employee or contract worker bears no significant relationship to the settled religious convictions of the organisation. Religious associations claiming the right to discriminate on otherwise illegal grounds should be required to show a connection between the position in respect of which they wish to discriminate and their beliefs and convictions. Contrary to the majority of the US Supreme Court in *Amos*, courts should not simply suspend their own judgment and defer to the definition offered by religious associations of what activities and jobs are sufficiently close to their religious beliefs. They should scrutinise the religious basis of claims to be permitted to discriminate on a case-by-case basis.

On the spectrum of greater to lesser proximity to the doctrinal core of the religion, we have at one end spiritual and religious leaders, in respect of whom a religious association should be permitted to discriminate on otherwise prohibited grounds. At the other end of the spectrum are the activities of ‘ordinary’ employees and contract workers such as secretaries and janitors which are too distant from the doctrinal core of the religion to justify their inclusion under an exemption. There are, however, positions

---

145 Galston (n 40 above) 185.
146 *R* (n 117 above) para 44.
intermediate between these two poles: A teacher whose duties extend only to non-religious subjects occupies one such a position. It may be difficult, admittedly, to judge when such intermediate positions are sufficiently close to matters of faith to mandate a constitutional exemption, but such a determination is necessary in each case.

Where legislation prohibiting discrimination subjects a religious institution to a substantial burden (to be determined in the circumstances of each case) by interfering with religious convictions and purposes, this provides a strong reason to permit such organisations to discriminate. But this reason must be balanced against whatever burden is placed on the employee or contract worker (or potential employee or contract worker) as a victim of discrimination. The extent of the burden will often depend on whether alternative work opportunities are available. If the burden placed on the individual is insubstantial and that placed on the religious group is severe, the liberal commitment to non-discrimination may have to yield.
ON THE FRAGILITY OF ASSOCIATIONAL LIFE: A CONSTITUTIVE LIBERAL’S RESPONSE TO PATRICK LENTA

Stu Woolman*

You wouldn't want to live in a world where you can't be conned, because if you were, you would be living in a world with no trust. That's the price you pay for trust – being conned.

Ricky Jay and Errol Morris

1 Introduction

1.1 The facts and the ‘new’ facts

With his customary wit, elan, and analytical rigor, Lenta's article 'Taking diversity seriously' has (a) critiqued and reconstructed the Equality Court's judgment in *Strydom v Nederduitse Gereformeerde Gemeente Moreleta Park*,¹ and (b) in so doing, has put our equality/association jurisprudence on a somewhat more solid footing. Because Lenta and I share a methodological predisposition towards the analysis of such cases, it should come as no surprise that we cover similar philosophical terrain and find ourselves roughly in agreement.²

² Lenta states that the primary ‘purpose [of his article] is to determine the appropriate domain of anti-discrimination law insofar as government wishes to prevent work-related discrimination by, and within, religious associations. Claims by religious associations that they should be permitted to engage in work-related discrimination on otherwise illegal grounds give rise to a clash of rights: the rights to freedom of association and freedom of religion on the one hand, and the right to equality on the other. My aim is to ascertain, with reference to *Strydom*, the circumstances in which it is
Did I say roughly? Yes, roughly.

For reasons that I will make immediately clear, rough agreement is the best that we are going to do with respect to the various arguments in this matter brought to bear by the parties to Strydom, the Equality Court, Professor Lenta and me. The problem, however, is not with the arguments in Strydom. It is with the facts. For what seems at first, second and third blush to be a conflict between equality rights, on one hand, and associational rights, religious rights and community rights, on the other, is actually nothing of the sort.

I believe that in our haste to prepare papers for our colloquy, Lenta and I failed to realise the actual nature of the matter that seized the High Court. As a result, I am now inclined to say that nothing of genuine constitutional import was at stake in Strydom.

Here are the facts of Strydom, as Lenta pithily puts them:

acceptable to privilege the rights to religious liberty and freedom of association over the right to equality: circumstances, that is, in which religious associations should be permitted to engage in work-related discrimination on otherwise prohibited grounds. Lenta (n 1 above) 78. His conclusion runs, in brief, that 'the Equality Court's reasoning in Strydom – including, principally, its privileging of equality over freedom of religion and associational liberty, and its trivialising of the burden placed on the church by continuing with Strydom's services – reflects a failure to take diversity sufficiently seriously, even if its conclusion is correct. Taking diversity seriously entails that the state does not have the right to abolish unfair discrimination in all its forms wherever it might appear: The right of a liberal state to enforce public principles such as non-discrimination is limited by the rights to freedom of association and freedom of religion. Latitude should be accorded to religious associations to allow them to govern their internal affairs and that includes accommodating their otherwise illegal work-related discrimination. At the same time, a religious association does not have the right to discriminate on otherwise prohibited grounds in respect of all activities performed by its employees and contract workers. Religious bodies should not be permitted to engage in work-related discrimination where the activity to be performed by the employee or contract worker or prospective employee or contract worker bears no significant relationship to the settled religious convictions of the organisation. Religious associations claiming the right to discriminate on otherwise illegal grounds should be required to show a connection between the position in respect of which they wish to discriminate and their beliefs and convictions. [C]ourts ... should scrutinise the religious basis of claims to be permitted to discriminate on a case-by-case basis.' Lenta (n 1 above) 108. I had previously described this article as an 'amplification' of Lenta's work. And that it is – to the extent that Lenta does not press down hard enough in attempting to describe the appropriate and the inappropriate domains for fair discrimination in religious associations. His casuistic approach – a case-by-case basis for analysis – does not do the subject matter justice nor give readers sufficient guidance. Yet I remain steadfast in my conviction that Prof Lenta and I occupy quite proximate positions within the very catholic 'liberal' political tradition of constitutional lawyers – in direct opposition to referees who would depict our respective 'liberalisms' otherwise. Admittedly, while our proximity actually gives our differences in approach their bite, it might lead some readers to view the exchange as a 'dust-up'. I think, to the contrary, that our colloquy is best described as an opportunity to learn aggressively from one another.

On the fragility of associational life: A constitutive liberal’s response to Patrick Lenta

The Equality Court had to determine whether a religious association should be granted an exemption from the Promotion of Equality and Prevention of Unfair Discrimination Act to permit it to engage in work-related discrimination on the basis of sexual orientation with respect to a music teacher. The issue confronting the Court was whether a religious association’s right to discriminate on prohibited grounds should extend beyond religious leaders to include a music teacher in a faith school not involved in religious instruction.4

The Equality Court held that it should not and that the actions of the church constituted unfair discrimination in terms of the Promotion of Equality and Prevention of Unfair Discrimination Act (PEPUDA).5

You can see the problem, can’t you? On the facts of the matter, the Strydom court reduced religious, associational and community rights regarding ‘membership, voice and exit’ to the ability to hire and to fire a ‘religious leader’.

Strydom’s facts, on closer inspection, suggest that:

(a) nothing that rises to the level of constitutional solicitude for religious belief or religious communal practice was at stake for the community;
(b) nothing that rises to the level of constitutional solicitude for religious belief or religious communal practice was at stake for the music teacher;6
(c) the music teacher could well have taught elsewhere since the religious character of the school was not an essential part of the job; and because the religious character was not an essential part of the job for her and she could have sought work elsewhere, then her dignity was not impaired.

If Strydom’s facts are so, then the dispute is not about religious, associational or community constitutional rights and how they ‘man up’ against rights to dignity and equality. Strydom is a straight up and down PEPUDA dispute that rightly rejects sexual orientation as a legitimate ground for termination of employment where no persuasive justification can be offered in rebuttal. The point that I wish to make appears in the formulation of Lenta’s proposition 8, namely that the nexus between the work of this organist and the ‘specific’ tenets of this church alone are not sufficiently close. Well then, if the respective interests are not sufficiently close, then there is, to put it mildly, no conflict of constitutional rights and no truly interesting constitutional question engaged in Strydom.

Now, to make matters more interesting, let us alter the facts slightly so that we might in fact learn something about how rights to religious

---

4 Lenta (n 1 above) 78.
5 Act 4 of 2000.
6 Even if it is obvious, on the facts, that the music teacher’s associational rights or religious rights were not at issue, it is worth emphasising this fact. Why? Cases with genuine purchase generally pit conflicting associational and religious rights against egalitarian claims.
practice, rights to association and rights to community 'man up' against rights to dignity and equality. In short, let us construct a hypothetical case (a) in which something is truly and palpably at stake and (b) our respective liberal views regarding how such conflicts ought to be resolved actually do some work.

Here are the facts of a high-stakes constitutional matter:

(a) Choir participation or organ playing at Catholic mass requires membership in the community – and the taking of communion.

(b) It was clear from any voluntary or commercial arrangement struck that the church organist understood the rules of the religious community and accepted that adherence to those rules was essential for any ongoing voluntary or commercial transaction that involved religious rites.

(c) The organist played mass on Saturday night and Sunday knowing full well that her sexual orientation conflicted with basic tenets of the religious community. (Indeed, under the church's 'don't ask, don't tell policy', the priest, the organist and many members of the community aware of her orientation decided to keep quiet.)

(d) Only once the organist's sexual orientation was brought to the attention of the church elders, and a formal inquiry was undertaken by a properly-constituted church board, did the church take the painful, but what it deemed religiously requisite act, of discharging the organist.

(e) However, let us first acknowledge that the church board first contacted the local bishop, who, sympathetic to the situation, contacted a cardinal in Rome; the cardinal, somewhat vexed by the problem, decided that he did not have the authority to take a final decision; he, in turn, contacted the Pope.

(f) The Pope, after contacting counsel, decided that while one could remain silent about one's sexuality within the church and still belong to and participate in the community, church scripture and doctrine made it clear that membership within the community turned on adherence to clearly delineated rules that made public acknowledgment of lesbianism and homosexual activity a grounds for removal from any church office or from participation in the rights of any parish.

(g) The Pope noted that this conclusion greatly pained him and that he hoped that the organist would reconsider her sexual orientation and return to the church as a fully-practising member of the parish.

Now we have an argument that liberals – of varying stripes – should find worth fighting about: the (largely infallible) doctrinal pronouncements of the Pope and the potentially conflicting dictates of a basic law committed to the rights to dignity, equality, association and religious belief and practice.7

7 Several commentators have mistakenly construed my views about this 'hard' case for my views about the outcome in Strydom. For a fuller account of my views in such matters, see S Woolman 'Seek justice elsewhere: An egalitarian pluralist reply to David Bilchitz on the distinction between differentiation and domination' (2012) 28 SAJHR 273.
1.2 The holding and the 'new' holding

It should come as no surprise that I am not so certain that Strydom is correctly decided (even when reconstructed by Lenta). Indeed, had the church provided a well-constructed argument grounded in (1) freedom of association (section 18); (2) freedom of religion (section 15); and (3) freedom of religious community practice (section 31), I would, given the absence of a compelling countervailing dignity claim by Strydom, have held in favour of the Nederduitsche Gereformeerde Gemeente Moreleta Park. For, as morally repugnant as I think that their beliefs and practices may be, it is important to remember that moral repugnance is not the same thing as, and does not always map directly onto, a finding of constitutional infirmity.

1.3 Structure of the argument

This reply to Lenta takes the following form: After acknowledging that we are both liberals of slightly different stripes, I go about demonstrating exactly what kind of liberal I am. To the extent that (my) liberalism, like any other liberal political theory, requires a foundation, I ground my version in a constellation of related notions: the constitutive nature of the self and the social; the need to protect the literal and the figurative property of associations from capture in a liberal society; the social capital that makes virtually every setting for meaningful action possible; and the need to leverage existing stocks of social capital so as to enable historically disadvantaged individuals and groups to pursue lives that they have reason to value.8 Having set out the foundation for constitutive liberalism, I turn to the question at hand: What must my kind of liberal do when confronted by competing constitutional claims by an individual excluded from meaningful participation in a religious community because of non-compliance with that community's rules, on the one hand, and by a community that contends that the ability to live by and to enforce the rules that define it is the only manner in which that community qua community can survive? By answering this question, I can then offer an assessment of the outcome in Strydom and my own, more difficult, hypothetical matter.9

---

8 See A Sen Development as freedom (1999).
9 Before I press on, I want to pause to note that, while Prof Lenta and I work within the same broad church, the Anglo-American political-constitutional tradition, we do not work that tradition in exactly the same way. Prof Lenta and I both agree that the Anglo-American political-constitutional tradition truly begins with John Locke's A letter of toleration (1689). But the meaning of that 'letter' is not altogether clear. Locke's primary concern – on my reading – was finding a political-philosophical basis for a negotiated settlement that would prevent England from being continually riven by religious strife. The primary reason the First Amendment (1791) is the first amendment to the US Constitution is freedom of religion – and not, as might be commonly thought, freedom of expression. Most Americans of European dissent had fled their native lands in order to escape religious persecution. Alighting on American soil, many of the denizens of the 13 colonies set about producing powerful local and colonial governments committed to
2 Constitutive attachments, capture and social capital

2.1 Constitutive attachments

For my kind of liberal: 'Meaning makes us.' That bumper sticker flows from one basic proposition. There is rarely a thought in your head or an 'intentional' or unintentional action that you undertake that is not directly sourced in or arising from pre-existing cognitive routines, dispositional states or collective practices. Let's break that complex proposition into two smaller parts.

2.1.1 The self and the constitutive

We, as a species, but also within the Western philosophical tradition (and I am not knocking either), tend to overemphasise dramatically the actual space for self-defining choices. In truth, our experience of personhood, of self, is a function of a complex set of narratives over which we exercise little in the way of (self-) control. Our notion of 'selfness' is a function, a very useful by-product, of a complex array of semi-independent neural networks that control the body's journey through life. This complex set of dispositional states is a function of both the deep grammar of our brains and the social endowments that have evolved over time to determine various patterns of behaviour. It should be apparent from this brief account that the self is a valuable abstraction and not an entity that stands back from experience and then dictates to the body what it does in response to various stimuli. Each self then is just a centre of narrative gravity. Each self is, without question, unique. (That uniqueness flows from obvious differences in biological make-up and social construction.) The variety of narratives that make up 'me' is different in a sufficiently large number of respects to allow me to differentiate my 'self' from any other 'self'. This 'self' is relatively stable. Though my narratives and dispositional states are always changing, my self-representations enable me to see my 'self' as advancing their particular comprehensive vision of the good (as reflected by the tenets of their particular Christian faith). The genius of the First Amendment of the US Constitution lies not in some desiccated negative liberal theory of the polity that many attribute to it. Its radical break with any previous founding political document is reflected in a uniquely American solution to the problem of how communities with vastly different conceptions of the good could live peaceably amongst each other via a document that provided a 'theory' of a just political order that satisfied all 13 states. It is this notion that a liberal constitutional order – as we have in South Africa – is primarily committed to enabling communities with vastly different conceptions of the good to live peaceably amongst one another is what drives this article. Here we have, on full display, in both the South African Constitution and the US Constitution, the genius of liberalism: the ability to negotiate between the right and the good.

10 For more on the self as a centre of narrative gravity, see D Dennett Consciousness explained (1991) 167-171. See also S Woolman The selfless Constitution: Experimentalism and flourishing as foundations for South Africa's basic law (2012).
remaining relatively consistent over time. But again, keep in mind that the self, and its various narratives, is thoroughly a function of physical capacities and social practices over which I have little control or choice.

2.1.2 The social and the constitutive

We also tend to overemphasise dramatically the actual space for rational collective deliberation. We often speak of the associations that give our lives meaning and content as if we were largely free to choose them or make them up as we go along. I have suggested why such a notion of choice is not true of us as individual selves. It is also largely not true of associational life generally. As Walzer has argued, there is a 'radical givenness to our associational life'.11 What he means, in short, is that most of the associations that make up our associational life are involuntary associations. We do not choose our family. We generally do not choose our race or religion or ethnicity or nationality or class or citizenship. Moreover, even when we appear to have the space to exercise choice, we rarely create the associations available to us. The vast majority of our associations are already there, culturally determined entities that pre-date our existence or, at the very least, our recognition of the need for them. Finally, even when we overcome inertia and do create some new association (and let me not be understood to underestimate the value of such overcoming), the very structure and style of the association is almost invariably based upon an existing rubric. Corporations, marriages, co-edited and co-authored publications are modelled upon existing associational forms. So gay marriages may be a relatively new legal construct, but marriage itself is a publicly-recognised and sanctioned institution for carrying on intimate or familial relationships. Even in times of transformation and revolution, reiteration and mimicry of existing associational forms are the norm.

Lenta tries out a different angle on the problem of associational freedom:

Another way to capture the significance of freedom of association is to frame it as a matter of diversity. As Galston has argued, "properly understood, liberalism is about the protection of diversity".12 A liberal state is one in which there exist associations embodying contrasting conceptions of the good life and differing views on what constitute worthy goals. As Brennan J puts it in *Jaycees*, associations "foster diversity".13

---

However, Galston, Brennan and Lenta all beg the question as to why we should value diversity – especially where it runs counter to claims made by historically-disadvantaged groups. The answer to that question does not lie in some intrinsic feature of ‘diversity’ but in two quite important facets of a truly ‘diverse’ polity: (a) A person must actually have the ability to exit the traditional (and sometimes quite closed and repressive) community (within the polity) into which she or he was born or raised; and (b) we, as members of various communities and non-members of others, must have the ability to experiment and to see what communities ‘fit’ or ‘work’ best for us.

2.1.3 Constitutive attachments and the setting for meaningful action

But before we proceed to ‘exit’ and ‘fit’, it is important to remember that a constitutive liberal, such as myself, emphasises rights to religious freedom, freedom of association and rights to religious practice because various forms of human association provide goods in some sense independent of their dominant feature (religion, trade culture, etc). Again: (1) Associations, such as religious communities, are integral to self-understanding, and that there can be no self without the host of associations that give the self content; and (2) associations, such as religious communities, enhance social cohesion, and are the indispensable settings for meaningful action. This two-fold recognition must be recast as two important questions for courts that evaluate the importance of associational rules and religious practices that bump up against other important constitutional considerations: (a) Is the association in question important for individual and group identity? (b) Is the association the setting for meaningful action? Here, the deep background considerations that a court should keep in mind when considering challenges to associational freedom are two-fold. First, the court should be quite wary of supplanting an association’s preferred vision of the good with one of its own or that of a democratically-elected majority (or even a powerful, wealthy, legally-savvy minority). Constitutions are designed to protect a significant degree of private ordering. Again, some commentators might prefer a radically egalitarian centrally planned political order. The South African Constitution’s framers opted for a decidedly social democratic order consistent with pronounced heterogeneity of its constituent sub-publics. Second, and perhaps more importantly, a court must take great care before it opens up, interferes with or dismantles existing associations. Such interventions must be carefully calibrated to preserve existing social capital at the same time that the imperative of transformation grants access to that capital and its concomitant opportunities.

Lenta appears to agree – at least partially – with the aforementioned position when he writes: ‘[O]ne cannot conceive of meaningful individual religious freedom unless religious institutions are protected from
government encroachment. Indeed, once the interests protected by the right to freedom of religion are added to those protected by the right to associational liberty it may be that, as Greenawalt contends, claims by religious associations for exemption from anti-discrimination laws 'will often have more force than claims deriving from other associations'.

Neither freedom of religion nor associational liberty, however, are absolute rights nor is the right to equality an absolute right. Freedom of association and religious liberty will sometimes have to give way to the right to freedom from discrimination.

However, that last sentence gives him away, at least in so far as what separates us. I want to claim that rights to be free from discrimination often must give way to freedom of religion and freedom of association if we are to maintain innumerable social settings for meaningful action.

2.2 Capture

Concerns about what I call 'capture' lie, like 'constitutive attachments', at the very heart of associational life. Indeed, concerns about capture are, essentially, a function of – one might even say a necessary and logical consequence of – the very structure of associational life.

In short, 'capture' refers to and justifies the ability of associations to control their association through selective membership policies, the manner in which they order their internal affairs and the discharge of members or users. Without the capacity to police their membership policies, as well as their internal affairs, associations would face two related threats. First, an association would be at risk of having its aims substantially altered. To the extent the original or the current raison d'être of the association matters to the extent members of the association, the association must possess the ability to regulate the entrance, voice and exit of members. Without built-in limitations on the process of determining the ends of the association, new members, existing members and even outside parties could easily distort the purpose, the character and the function of the association. Second, an association's very existence could be at risk. Individuals, other groups or a state inimical to the beliefs and practices of a given association could use ease of entrance into and the exercise of voice in an association to put that same association out of business. In sum, by 'capture', I mean that in order for most associations to function as associations, they must possess a substantial degree of control over who belongs to the association and some degree of control over the ends the association pursues. So long as the association, and its members, as

16 Lenta (n 1 above) 82.
currently constituted possesses a figurative and/or real sense of ownership, so long as there is real social capital at stake, a court (and the state) must cede to the religious association a significant level of control over entrance, voice and exit.

Does Lenta ever seriously grapple with capture? I do not think so. And to the extent that he does, he certainly does not give it much credence:

In Bob Jones University, the Supreme Court does not rule that a private, religious university may be forced to abandon its internal, racially discriminatory policy on pain of civil or criminal sanction. Such a step, as Galston observes, would have implicated basic rights of freedom of association and free exercise of religion that the Court has long defined and protected, and it would have called for an entirely different legal analysis. Even employing the line of reasoning it used to settle Bob Jones, the Court would almost certainly have found that unlike the mere revocation of tax exemption, a direct order to change the discriminatory policy would have failed the balancing test by imposing an “undue burden” on petitioners’ free exercise.17

So the approach followed in Bob Jones University is on closer inspection hardly supportive of the approach taken in Strydom. Extending the precedent of Bob Jones University to the facts of Strydom would mean not invalidating the church's discrimination on a prohibited ground, but burdening the church by removing all forms of otherwise applicable public encouragement.18 Lenta fails to acknowledge the extent to which the elimination of state subsidies or exemptions may work the same black magic on a religion as direct assaults on the citadel.19

2.3 Social capital

One can identify a third, golden, thread running through the two previous justifications for associational freedom and collective religious practice. This leitmotif might best be described as social capital. Social capital is – and is a function of – our collective effort to build and to fortify the things that matter. It is our collective grit and elbow grease, our relationships and their constantly re-affirmed vows. Social capital emphasises the extent to which our capacity to do anything is contingent upon the creation and the maintenance of forms of association which provide both the tools and the setting for meaningful action. Social capital is often treated as ephemera. That makes sense. It is so hard to see. In fact, it is this elusive quality that makes social capital so fragile. It is made up, after all, not of bricks and

17 Galston (n 12 above) 183.
18 Lenta (n 1 above) 89.
19 Internal capture is not the only danger. External capture poses an equally dangerous threat. Should elected elites of a conservative strain wish to reverse progressive decisions of the bench, it is easy enough to pack the courts with like-minded judges who would reverse the trajectory of a Constitutional Court's jurisprudence.
mortar, but of relationships and commitments, and the trust, respect and loyalty upon which they are dependent.

Lin characterises social capital as follows:

[S]ocial capital may be defined operationally as the resources embedded in social networks accessed and used by actors for actions. Thus, the concept has two important components: (1) It represents resources embedded in social relations rather than individuals, and (2) access and use of such resources reside with actors. The first characterisation, socially embedded resources, allows a parallel analysis between social capital and other forms of capital … For example, human capital, as envisioned by economists, represents investment on the part of individuals to acquire certain skills … that are useful in certain markets … The second component of social capital … must reflect that [the] ego is cognitively aware of the presence of such resources in her or his relations and networks and makes a choice in invoking the particular resources.20

In somewhat simpler prose, Halpern offers the following gloss:

So what is social capital comprised of? Most forms … consist of a network; a cluster of norms, values and expectancies that are shared by group members; and sanctions – punishments and rewards – that help to maintain the norms and the networks … The first component is the social network … in some cases the small rural village. The network can be further characterised by its density … and its closure … The second component is the social norms. These are the rules, values and expectancies that characterise the community of network members … Many of these rules are unwritten. The third component is sanctions. Sanctions are not just formal – such as punishments for breaking the law. Most are very informal, but are nonetheless effective in maintaining social norms … The sanction may be through someone being told directly … More commonly, however, the sanction is indirect and subtle, such as through gossip and reputation.21

20 N Lin Social capital: A theory of social structure and action (2002) 24-25. Social capital’s more famous expositors began writing about the phenomenon over two decades ago – and often attributed different meanings to the term. Bourdieu defines ‘social capital’ as made up of social obligations or connections … [and the aggregation of] actual or potential resources which are linked to possession of a durable network of institutionalised relationships of mutual acquaintance and recognition. See P Bourdieu ‘Forms of capital’ in JG Richards (ed) Handbook of theory and research for the sociology of education (1986) 241-248. However, only several pages later, he states in quasi-Marxist language that ‘economic capital is at the root of all the other types of capital’ and that ‘every type of capital is reducible in the last analysis to economic capital’. Bourdieu (above) 241-252-253. On the other side of the political spectrum, Putnam defines social capital as follows: ‘Whereas physical capital refers to physical objects and human capital to the properties of individuals, social capital refers to social networks and the norms of reciprocity and trustworthiness that arise from them.’ See R Putnam Bowling alone: The collapse and revival of American community (2000) 19. See also R Putnam Making democracy work: Civic traditions in modern Italy (1993). Putnam does offer a better definition of social capital elsewhere: Social capital consists of ‘networks, norms and trust – that enable participants to act more effectively together to pursue shared objectives … Social capital, in short, refers to social connections and the attendant norms and trust.’ See R Putnam ‘Tuning in, tuning out: The strange disappearance of social capital in America’ (1995) 28 Political Science and Politics 1-20.

The efforts of Halpern, Lin and Putnam have led the avatar of capitalism, the World Bank, to acknowledge the centrality of social capital:

Social capital refers to the institutions, relationships, and norms that shape the quality … of a society's … interactions. Increasing evidence shows that social cohesion is critical for societies to prosper and for development to be sustainable. Social capital is not just the sums of the institutions that underpin a society – it is the glue that holds them together.22

In light of Putnam, Lin and Halpern’s definitions, social capital can be understood to link together my primary justifications for the protection of associational life and religious practice in the following manner. Social capital is what keeps our intimate, economic, political, cultural, traditional, reformist and religious associations going. Without it, nothing works. Social capital explains at least part of what is at stake for both individual identity and social cohesion: the constitutive. Social capital recognises that we store the better part of our meaning in fundamentally involuntary associations. Squander that social capital, nothing that truly matters will continue to exist. Social capital recognises the dominant rationales for ceding control over membership and purpose to the association. Social capital recognises both the real and the figurative sense of ownership that animates associational life. If anyone and everyone can claim ownership of and in an association, then no one owns it. Social capital takes seriously the threat of various kinds of compelled association. Trust, respect and loyalty have no meaning where the association is coerced. These several virtues can be earned, but never commanded. No trust, respect or loyalty: no social capital.23 No social capital: none but the most debased association.24

2.3.1 Social capital: Bonding networks and bridging networks

But just as not all associations are alike – not all forms of social capital are fungible. For the purposes of this article, two forms of social capital networks are of particular import: bonding and bridging. Putnam puts the difference between these two distinct forms of social capital or social networks as follows:


24 Of course, associations have their dark, hierarchical and punitive dimensions. They may prevent members from securing the goods necessary to flourish. See S Woolman ‘Freedom of association’ in S Woolman et al (eds) Constitutional law of South Africa (2nd ed, 2003) ch 44. As I have explored elsewhere, one remedy that does threaten associational freedom might be a finding that an association that limits the life opportunities of a class of its members should, along with the state, provide the means for exit that would allow individuals to seek justice elsewhere. See S Woolman (n 7 above).
Some forms of capital are, by choice or necessity, inward looking and tend to reinforce exclusive identities and homogeneous groups. Examples of bonding social capital include ethnic fraternal organisations, church-based women's reading groups, and fashionable country clubs. Other networks are outward looking and encompass people across diverse social cleavages. Examples of bridging social capital include the civil rights movement, many youth service groups, and ecumenical religious organisations … Bonding social capital provides a kind of sociological superglue whereas bridging social capital provides a sociological WD-40.25

One way to distinguish further between the two networks would be to 'contrast the strong bonds of reciprocity and care that are found inside families and small communities (what we might call normative bonding social capital) with the … self-interested norms that tend to predominate between relative strangers … and through which relative strangers can cooperate successfully (what we might call 'normative bridging' social capital').26 But that is just a start. High-bonding communities tend to feature well-established, historically-entrenched belief sets, shared assets and rather rigid rules regarding membership, voice and exit (and enforcement mechanisms regarding those rules). Bridging networks are often extra-communal and bring together rather diverse groups of individuals in the pursuit of singular, generally self-interested ends. Membership, voice and exit tend to be both more flexible and more formal in bridging networks.

However distinct these two kinds of social capital networks may appear on the surface, I would argue that the success of a developmental state such as South Africa depends upon (a) respect for the significant public goods created by private bonding networks (schools, hospitals, charities); (b) leveraging, where possible, admission into bonding networks for persons (and groups of persons) who would otherwise not have access to the goods made available within those networks; and (c) the use of state resources to build linking or bridging networks that, over time, produce social capital that is comparable in nature and quality to that social capital produced in bonding networks.

2.3.2 Social capital: The fragility of associational life in South Africa

As De Tocqueville was first to note, medium to large-scale democracies only flourish in an environment with a rich associational life.27 Social theorists have come to recognise that there is a direct correlation between (a) the availability (or the lack of availability) of social capital; (b) the presence (or absence) of bonding networks and bridging networks; and (c) the virility (or the sterility) of political life. This syllogism is not the product
of armchair philosophers. As Halpern and colleagues have demonstrated, Nordic countries – such as Sweden, Norway and Denmark – often possess social capital or mutual trust ratios of over 65 per cent. That means, in short, that two-thirds of the citizens of these countries tend to 'trust' 'strangers' in their midst. Such trust enables them to work effectively with their fellow citizens. Even the United States, the land of the free, autonomous (fragmented) individual, boasts social capital or mutual trust ratios of over 50 per cent. South Africa posts a dismal 15 per cent Brazil – with an even longer history of social strife – is one of the few democracies to advertise a lower mark: 2 per cent.

South African political scientist Ivor Chipkin has recently argued that social cohesion is essential for long-term economic stability in South Africa. Building on recent debates about the importance of various kinds of social capital (especially bridging networks), Chipkin contends that state institutions – especially in a developing democracy such as South Africa – have an essential role to play in building social capital and promoting social cohesion:

There are several ways in which such linking is achieved. It may be that churches and other religious organisations, working on the basis of charity, are the key linking mechanisms between poor and resource rich(er) communities. Various civil-society bodies, including non-governmental organisations, may play similar roles. Yet the most important institution, in this regard, is the state. This is true for several reasons. In the first place, democratic state institutions, like local governments, are able to realise benefits, not simply for members of ascriptive groups, but for communities of citizens – irrespective of religious affiliation or culture or ethnicity. What matters is the degree to which their operations are inclusive and participatory and the degree to which they are able to invest in and/or leverage resources for poor communities ... In the second place, the democratic state builds networks and creates linkages on the basis of democratic values. In other words, they encourage a culture of democratic citizenship in the country.

In sum, the state has a critical role to play in ensuring that the associational life of our extremely heterogeneous society buttresses the egalitarian goals, the utilitarian interests and the democratic ends of our polity. It can do so through economic policies – micro-financing or black economic empowerment (BEE). It can do so through school admission and funding policies as well as the composition of school governing bodies. It can do so by ensuring that citizens are given a meaningful opportunity to participate

in the decision-making processes that have a direct impact on their lives.33

3  The over-valorisation of equality

One clearly-justifiable form of state interference with associations and networks is the requirement that certain associations open themselves up to a wider potential membership because they control access to important (and otherwise unavailable) social goods. PEPUDA does just that.34

But the Act goes further. The Act makes equality per se a trump. The Act presents a particularly difficult question for Lenta and me: When, if ever, does the freedom to associate or the right to religious practice secure a legitimate right to discriminate?

The US Supreme Court employs a set of doctrines that rests upon distinctions between 'expressive associations' and 'intimate associations', on the one hand, and those associations which are neither expressive nor intimate, on the other. Where an association is neither expressive nor intimate, the right to dissociate and discriminate may fall before some other important state interest.

These doctrines are articulated in a trio of cases: Roberts v United States Jaycees; Board of Directors of Rotary International v Rotary Club of Duarte; and New York State Club Association v City of New York.35 In all three public accommodation cases, constitutional associational interests fell before egalitarian concerns reflected in state statutes. But they only tell one side of the story. The other side is told in Boy Scouts of America v Dale.36

In Dale, the Boy Scouts of America have had their desire to exclude homosexuals vindicated by the US Supreme Court. A majority of the US Supreme Court justices found a relatively clear nexus between the right of the Boy Scouts of America to represent themselves as standing for a particular set of values (whether they actually expressed them or not) and the right to exclude individuals whose behaviour apparently subverts those stated values. Although not employed by the Court, the following three-part analysis suggested herein does some work in explaining what was at stake in Boy Scouts of America v Dale.

The Boy Scouts of America (a very large and powerful bonding and bridging network) offers a suitably complex setting to test some of my interim conclusions about bonding, bridging and association analysis. The

33 Doctors for Life International v Speaker of the National Assembly 2006 6 SA 416 (CC); Matatiele Municipality & Others v President of South Africa & Others 2007 6 SA 477 (CC).
34 n 5 above.
Boy Scouts of America pursues particular goals – the pursuit of certain virtues by young men. It distributes various goods – opportunities to learn skills, to participate in organised events and to secure a certain status within the broader community. Given these goals and goods, should the Boy Scouts of America be free to include and exclude whomever they like? As our 'functional' analysis suggests, the Scouts' freedom to exclude individuals raises three related concerns. First, such rules may discriminate against individuals and groups in ways wholly unrelated to the ends of the association. Second, such rules may discriminate against individuals and groups in ways that offend the rules to which the larger polity is committed. Third, a rigid adherence to existing articles or by-laws of an association may preclude the association from a natural or spontaneous evolution into an organisation that pursued a somewhat modified or a dramatically different set of ends. The first concern links the gay scoutmaster's exclusion to the association's ends. The second concern engages the issue of privileging the constitutive nature (and meaning) of the Scouts to (a majority of) its members over egalitarian concerns of the society in general. The third concern determines the extent to which the Scouts' governing authority may declare, through fiat, exclusionary criteria and what the ends of the association are.

While the US Supreme Court has been forced to grasp the nettle of membership practices, it is relatively clear from the outcome in Dale that it has not been willing to press down particularly hard on the Boy Scouts' stated justifications for exclusion. It is not clear how, if at all, the sexual orientation of a scoutmaster (and former Eagle Scout) runs counter to the general ethos of the Scouts. Likewise, it is not clear that the vast majority of members believe that the virtues promoted by the Boy Scouts are in any manner undermined by the sexual orientation of a scout or a scoutmaster. It is difficult to understand why homosexual status simpliciter should be understood to constitute expression or advocacy. To suggest that it does possess such expressive content, and then to find the status/expression to be of secondary constitutional import, is to suggest that the label 'homosexual' 'is tantamount to a constitutionally prescribed symbol of inferiority'.

This 'functional analysis' – and the conclusions it generates about Dale – dovetails with Lenta's conclusions about and the decision of the Equality Court in Strydom. As with the gay scoutmaster in Dale, it is not clear how the sexual orientation of the organist in Strydom (a) runs counter to the general ethos of the church (as opposed to allegedly biblically proscribed basic tenets of the faith); (b) offends the majority of the members of the community; or (c) constitutes an express endorsement of a particular kind

---

of sexual orientation that might influence the behaviour of another member of the community.

But while those questions about the justification of the discrimination cause Lenta to pull up short, I believe – for reasons already identified above – that the discrimination in Strydom (as opposed to Dale), even if morally repugnant from some vantage point, is constitutionally permissible. What is the difference that makes a difference between Strydom and Dale? The Boy Scouts – the voluntary association whose membership policies were at issue in Dale – simply do not rely on a relatively static belief set that gives its associational setting meaning. To wit: (1) The Boy Scout troops around the United States split, almost evenly, as to whether the gay Eagle Scout who became scoutmaster should remain in his post: Nothing in the Boy Scout manual addressed anything but the virtues exemplified by Dale; (2) As to sex and sexual orientation, it is now, and was then, common knowledge that many scout organisations (around the world and in South Africa) had eliminated the gender divide without any attendant loss of meaning or purpose – girls and boys, young women and young men, were viewed as possessing the same capacity for achievement and for the demonstration of outstanding character.

Like it or not, religion is different. Time – sacred time – for religion is circular, not linear. Linear time – human time – is profane. Religion returns through holiday and ritual to the beginning: when God – of one stripe or another – made the world and laid down the law. Others – non-members – may find the religious beliefs and practices of a faith perversive. But rational distinctions – or the absence of rational distinctions – is hardly the point when an organised faith is put in the position of defending God's law – not their own. It is God's law that provides the setting for meaningful action for the members of the Nederduitse Gereformeerde Gemeente Moreleta Park. The notion that they could alter God's law is akin to idolatry. It confuses the sacred and the profane.

Because the membership policies and the organisation of internal affairs are often critical to an association's identity, one must be aware that laws which force a change in these policies may alter the essential character of that association. So if one believes that political pluralism, cultural diversity, individual autonomy and social upliftment may be threatened by forced changes in an association's membership criteria – and one is committed as Lenta is, to 'taking diversity seriously' – then one might want to give associations the power to police their boundaries and thereby prevent the capture of the association by individuals or groups who might wish to change the association's aims. That is, one may wish to give the individuals and groups whose lives are substantially shaped by the associations of which they are a part, or who invest significant resources and effort in the maintenance and the creation of particular kinds of enterprises, the power to police the membership of the organisation in order to ensure that it remains true to its founding tenets.
Chapter 4

The acid test then is: When, or under what conditions, is an association entitled to exercise its right to determine its membership criteria (largely) free from external intervention?

4 The case for fair discrimination in religious associations

Given that Lenta and I share so many common articles of liberal faith (and thus would likely reach similar conclusions even in hard cases), it is important here to set out, again, those areas of divergence that might result in different outcomes.

Lenta promotes diversity as an archetypal, or even the most basic liberal value. And if we liberals were to begin and end our commitment to liberalism with a restatement of Locke's *A letter on toleration*, then I might be inclined to agree that a strong commitment to diversity – designed to keep religious civil wars under wraps – would be the primary justification for a liberal conception of the state and the manner in which it parses the right and the good.

However, I believe that account is too thin – and indeed too inaccurate with regard to the human condition – to justify liberalism. What makes liberalism truly formidable as a political, moral and constitutional ideology is that it takes not diversity, but associational life, seriously.

4.1 The easy question dispatched

I do not for a moment wish to suggest that Lenta and I differ substantially upon the core tenets of a liberal constitutional order. That said, there exist extremely important points of difference (of departure and arrival) that flow in large part from the manner in which Lenta has chosen to frame questions of associational freedom and discrimination.

As a result, the set of questions this article is obliged to answer concerns the extent to which religious associations ought to be able to discriminate fairly – across a range of domains – in furtherance of their faith and communal practices. Lenta first narrows the domain of inquiry (along with the *Strydom* court) as to whether or not the rights of individuals who suffer work-related discrimination on prohibited grounds should always trump the interest of religious associations in discriminating in accordance with their religious beliefs. That question is a 'red herring'. For the answer is 'surely not' – as Lenta notes – lest we wish the state to re-write scripture however it so wishes. That kind of political power surely has no place in an open and democratic society based on human dignity, equality and freedom, especially one expressly committed to freedom of religion and religious practice. No. The question that Lenta sets for himself is
whether such associations may legitimately discriminate in the 'religious workplace', when the nature of the discrimination appears rather weakly connected to basic tenets of a faith. Again, that question too appears to be something of a 'red herring'. For by building in the notion of a 'weak' relationship to the basic tenets of a faith, to the question of whether a religious association's right to discriminate on prohibited grounds should extend beyond the 'protection' of religious leaders to include a music teacher in a faith school not involved in religious instruction, Lenta has already put a thumb on the scale in favour of virtually any claim of discrimination – no matter how superficial – and causes us to lose sight of the questions about associational freedom and equality that truly matter. But before we turn to the hard question I set out at the beginning – what ought the courts to do when religious beliefs of some moment engage genuine concerns about the dignity of a member of the same community – let us take Lenta's concerns seriously.

Lenta is certainly alive to the independent importance of freedom of association in a liberal constitutional order: Why is freedom of association so important? One reason is that groups and association contribute to their members' wellbeing. Freedom of association enables individuals to 'create and maintain intimate relationships of love and friendship' and is 'increasingly essential as a means of engaging in charity, commerce, industry, education, health care, residential life [and] religious practice'. Associational freedom is an essential part of individual freedom: associations represent the choices of their members about how to live. The value placed on freedom of association also reflects the importance of civil society, 'that network of intimate, expressive, and associational institutions that stand between the individual and the state', discussion of which in the US political tradition goes back to Alexis de Tocqueville.38 The existence of associations serves as a counterweight to potentially overweening state power.39

Note the difference of emphasis here. For Lenta, it is all about how associations service (or 'represent' (his word)) individuals – individual 'well-being', individual 'choice' and a buffer between the state and the individual. Such justifications are standard – and I accept them myself. But, on my account, the order of priority is reversed. Community comes first as the bearer of meaning. The difference is a standard part of liberal discourse. Let me just remark that it is the apparent 'voluntariness' of these associations ('you can always find another') that results in the lack of seriousness with which the right to freedom of association is often taken. Can you name a Constitutional Court decision that takes freedom of association seriously – let alone a decision that expressly decides the matter on associational grounds? No: You cannot. In 15 years, the

38 Galston (n 12 above) 531.
39 Lenta (n 1 above) 80.
Constitutional Court has never decided a case solely, or even largely, in terms of freedom of association. And that says quite a lot about where, in our hierarchy of rights, freedom of association is located.

Lenta supplements the 'buffer argument' with another rather modern liberal justification for associational freedom: diversity. Lenta writes:

Another way to capture the significance of freedom of association is to frame it as a matter of diversity. As Brennan J puts it in *Jaycees*, associations “foster diversity”. He expands on this point as follows: “An individual's freedom to speak [and] to worship … could not be vigorously protected from interference by the state unless a correlative freedom to engage in group effort toward those ends were not also guaranteed. According protection to collective effort on behalf of shared goals is especially important in preserving political and cultural diversity.”

We all 'love' diversity. Or do we? Frankly, I could care less if a plethora of associations or religious, cultural or linguistic 'forms of life' disappeared tomorrow. They do nothing for me today – and they shall do nothing for me over the course of a lifetime. Why should a host of alien ways of being in the world make diversity important to me? (The question has an answer – but it has nothing to do with rainbow nations or melting pots or cultural mosaics. Hint: It has everything to do with how we derive better answers as to how we should or can lead a 'good' life. Answer: Sometimes other 'ways of being in the world' – rather than the ways of being into which we are born – actually provide a better fit for who we are as individuals.)

Lenta flirts with taking association seriously when he writes:

Indeed, once the interests protected by the right to freedom of religion are added to those protected by the right to associational liberty it may be that, as Kent Greenawalt contends, claims by religious associations for exemption from anti-discrimination laws “will often have more force than claims deriving from other associations”.

But Lenta never quite buys Greenawalt's point. Indeed, Lenta's very next paragraph suggests that the mutual reinforcement of the two rights provides neither right – nor his argument – with any greater purchase:

Neither freedom of religion nor associational liberty, however, are absolute rights. Nor is the right to equality an absolute right. Freedom of association and religious liberty will sometimes have to give way to the right to freedom from discrimination.

One can hardly count that as an endorsement of Greenawalt's position – or any position that gives religious associations a leg up in any dispute

---

40 Lenta (n 1 above) 81.
41 Lenta (n 1 above) quoting Greenawalt (n 15 above) 136.
42 Lenta (n 1 above) 82.
which involves discrimination on the basis of communal relations. However, it is Lenta's subsequent treatment of *Amos v Corporation of the Presiding Bishop* that makes it patently clear that he has little patience for line drawing that might tip the balance of a discrimination dispute in favour of a religious association. Lenta writes:

In *Amos*, a janitor employed for 16 years at a Mormon-run gymnasium was dismissed for his failure to meet the requirements for a certificate of eligibility for attendance at Mormon temples. The dismissal clearly constituted job discrimination on the basis of religious affiliation, prohibited under Title VII. The janitor challenged the constitutionality of section 702 exemption inasmuch as it permitted religious associations to discriminate on religious grounds in hiring for non-religious positions, into which category his own job fell. The District Court ruled that the amendment to the Title VII exemption was unconstitutional because it enhanced the ability of religious associations to further their beliefs in situations where 'people's opportunity to earn a living is at stake'.

The District Court's ruling is consistent with Basson J's holding in *Strydom* that certain jobs (typists, for example) are so distant from the doctrinal core of the church's activities as to render an exemption from anti-discrimination legislation unwarranted. A unanimous US Supreme Court, however, declined to follow the District Court. In his concurring judgment, Brennan J observed that 'religious organisations have an interest in autonomy in ordering their religious affairs'. Brennan J … decided that … [it] would be inappropriate since it would involve 'ongoing government entanglement in religious affairs' which could result in a religious organisation being 'chilled in its free exercise activity'. He expands on this point as follows:

While a church may regard the conduct of certain functions as integral to its mission, a court may disagree. A religious organisation therefore would have an incentive to characterise as religious only those activities about which there likely would be no dispute, even if it genuinely believed that religious commitment was important in performing other tasks as well. As a result, the community's process of self-definition would be shaped in part by the prospects of litigation.

Brennan J concluded that the 'categorical exemption' contained in section 702 (as amended) covering all types of church employee, no matter what the job, appropriately balances the right of employees to religious liberty and the autonomy of religious organisations. In contradistinction to the

---

43 *Amos v Corporation of the Presiding Bishop* 594 F Supp 791, 818 (Utah 1984).
45 *Amos* (n 44 above) 343.
46 *Amos* (n 44 above) 343-344.
47 *Amos* (n 44 above) 345-346.
District Court, the Supreme Court upheld the constitutionality of section 702.48

Lenta contends that the Amos court is incorrect. And here we – he and I – definitively part company. Lenta apparently believes that it is okay for courts to delve (carefully) into questions of core tenets of belief – and decide which beliefs are trumps and which beliefs are not. Brennan is alive to that problem, and alerts his readers to the dangers of according the courts such powers. The first question is whether there is a real cost imposed on the party challenging the religious association's criteria for entrance or employment. Where children are involved, the courts are entitled to intervene – because children lack the power to stand up to parents or to other authorities. I would say that the same might be true of adults who lack the requisite power to stand up to authorities or, more importantly, lack the ability to exit from a community on largely their own terms. With children or adults who lack the capacity to protect their own interests, I would defend a dignity claim asserted on their behalf. But where the interest of the claimant is purely pecuniary, and other job opportunities truly exist elsewhere, I can see no genuinely compelling constitutional interest to buttress his claim. Indeed, that is exactly why the US Supreme Court finds as it does in Amos: The constitutional right to free exercise of religion – buttressed by the freedom to associate – is rightly deemed stronger than a right to a particular form of employment. Our Constitutional Court has followed a largely similar route.50

Now let us put the real problem properly on the table. The question is not one of religious discrimination. (Neither the janitor nor the organist is interested in religious inclusion.) It is a question of whether a religious association can control its own real and figurative property. Here is a partial answer: The link between religious associations, constitutive attachments, social capital and hard capital throws up a most complex set of problems. Religious communities often support financially other

48 Lenta (n 1 above) 91.
49 Christian Education of South Africa v Minister of Education 2000 4 SA 757 (CC) (dignity interests of children trump religious beliefs in virtue of corporal punishment); Bhe v Magistrate, Khayelitsha & Others 2005 1 SA 580 (CC) (dignity and equality interests of female children trump interest of traditional communities in rule of male primogeniture).
50 Ex Parte Gauteng Provincial Legislature: In Re Dispute Concerning the Constitutionality of Certain Provisions of the Gauteng School Education Bill of 1995 1996 3 SA 165 (CC). (The Court held that sec 32(c) of the interim Constitution permitted communities to create schools based upon common culture, language and religion. It further held that sec 32(c) of the interim Constitution provided a defensive right to persons who sought to establish such educational institutions and that it protected that right from invasion by the state.) As the High Court in Wittmann v Deutscher Schulverein, Pretoria & Others 1998 4 SA 423, 451 (T) 451 reasoned, the right to maintain a private German school educational institution based upon culture, language and religion is predicated upon the capacity to exclude non-speakers, non-believers or non-participants. ('Freedom of association entails the right with others to exclude nonconformists. It also includes the right to require those who join the association to conform with its principles and rules.') See also Taylor v Kurstag 2005 1 SA 362 (W) para 38.
members of their community: often with employment. (Should they be able to do so at the expense of other non-members? Yes: The favourable treatment strengthens the bonds of trust, loyalty and mutual respect that successful associational life requires.) Quite often the access to capital comes with terms extraordinarily favourable to the borrower. How should we treat these compacts? Such economic arrangements are easily subject to charges of bias. That goes without saying. The question is: What can we do about it? Neither outlawing such funds nor opening them up is likely to have the desired outcome. The fund either goes underground or out-of-existence. The aim of such funds is the strengthening of community life. To the extent that the fund continues to serve such an aim (one that ultimately promotes religious life), the grounds for overriding its discriminatory and exclusionary policies and aims are limited.51

Several possible approaches exist to test the constitutionality of the accretion of economic capital in religious collectives. First, one could permit the accumulation and re-distribution of such capital within a particular community subject to the proviso that such capital could not be used by community members to achieve either dominance or monopoly power with respect to the distribution of social goods – say political power – in non-economic domains.52 Second, according to Gutmann’s civic equality principle, the test for such religiously-based economic decisions would be whether they still permitted the state to discharge its obligation to provide for (1) equal citizenship or political participation; (2) equal liberty; and (3) equal, basic opportunity to live a preferred vision of the good.53 Neither the janitor in Amos nor the organist in Strydom is being denied any of these goods. I would be far more alive to constitutional challenges to traditional and religious practices (say female genital mutilation) that objectively diminish the life opportunities and experiences of many South Africans.

### 4.2 Answering the hard question

At the outset, I noted that Lenta set what he believed to be a relatively easy question for a liberal to answer. I hope that the foregoing pages have demonstrated that the easy question is not so easy, and that the answer given by Lenta and the Strydom court may well be wrong. Now I want to turn our attention to the harder question – where a claim for religious belief and community practice exists on both sides, and where the individual plaintiff can plausibly maintain that the association has engaged in a *prima facie* impairment to her right to dignity.

---

51 Here is the flaw in Lenta’s crystal bowl: He does not take sufficient account of the communitarian dimensions of our basic law.
52 See eg M Walzer *Spheres of justice* (1983).
The hard question turns, I contend, on a rift in liberal thought that would not – could not – have been apparent to Locke when he wrote *A Letter on Toleration*. What did Locke miss? What he missed is the difference between a politics of respect that issues from claims grounded in human dignity and a politics of difference that issues from claims grounded in equal recognition.

It may well be that, in many societies, individual liberty, multicultural recognition and nation building are incompatible. Indeed, for a society in transition, multicultural recognition and national identity formation appear to pull in opposite directions. For even if individual identities are formed in open dialogue, these identities are largely shaped and limited by a pre-determined set of scripts. Collective recognition becomes important, in large part, because the body politic has denied the members of some group the ability to form – on an individual basis – a positive identity. In a perfect world, the elimination of group-based barriers to social goods would free individuals to be whatever they wanted to be. But even in a perfect world, claims for group recognition do not dissipate so readily.

What is the basis for the demand for group recognition? In any multicultural society, two different kinds of claims for equal respect and two different senses of identity sit uncomfortably alongside one another. The first emerges from what Taylor calls a politics of equal dignity. It is based on the idea that each individual human being is equally worthy of respect. That version of equal respect is what Lenta admirably defends. The second issues from a politics of difference. This form of politics tends to revolve primarily around the claim that every group of people ought to have the right to form and to maintain its own – equally respected – community and its own equally respected vision of the good.

The important distinction between the two is this. The first focuses on what is the same in all of us – that we all have lives and hopes and dreams that we should all be allowed to pursue. The second focuses on a specific aspect of our identity – our membership in a group – and says that the purpose of our politics ought to be, ultimately, the nurturing or the fostering of that particularity. The power of this second form of liberal politics springs largely from its involuntary character – the sense that we have no capacity to choose this aspect of our identity. As I have argued earlier: This source of meaning chooses us; it makes us.

One of the problems South Africa faces today is that it is difficult, if not impossible, to accommodate both claims. As Taylor himself notes, while ‘it makes sense to demand as a matter of right that we approach … certain cultures with a presumption of their value … it cannot make sense to demand as a matter of right that we come up with a final concluding judgment that their value is great or equal to others’. But the demand for political recognition of distinct religious communities often amounts to that second demand. Moreover, such recognition often reinforces a
narcissism of minor difference that, in turn, provokes anxiety about the extent to which members of other groups secure access to the most important goods in a polity. Such anxiety about a just distribution of goods – and the manner in which group affiliation distorts that distribution – necessarily interferes with national identity formation. The African National Congress has, for both historical reasons and for reasons associated with its vision of transformation, refused to lend significant support to group politics. The Constitutional Court is also predisposed towards claims of equal respect (for individuals and, occasionally, groups) grounded in a politics of equal dignity.

Does that mean that we should be group identity sceptics? Sen contends as follows:

Our shared humanity gets savagely challenged when the manifold divisions in the world are unified into one allegedly dominant system of classification – in terms of religion, or community, or culture, or nation, or civilisation … The uniquely partitioned world is much more divisive than the universe of plural and diverse categories that shape the world in which we live. It goes not only against the old-fashioned belief that ‘we human beings are all much the same’ … but also against the less discussed but much more plausible understanding that we are diversely different. The hope of harmony in the contemporary world lies to a great extent in a clearer understanding of the pluralities of human identity, and in the appreciation that they cut across each other and work against a sharp separation along one single hardened line of impenetrable division.54

That much seems incontestable. Totalising views of identity (with their ostensibly comprehensive visions of the good) have led to a hardening of boundaries between groups. The hardening of boundaries has led, in turn, to a hardening of hearts that enables many nations (and many communities or groups with claims to nationhood) to pillage, bomb and plunder with increasingly greater abandon.

The more difficult question for group identity sceptics in South Africa is how to draw down on our constitutive attachments in a manner that both (a) protects the social capital that we require to build the many institutions that make us human; and (b) prevents specific religious, cultural, and linguistic communities from using that social capital to undermine our ‘more perfect union’. Here Sen knows he – or we – are in trouble, but can only state the problem thus:

The sense of identity can make an important contribution to the strength and the warmth of our relations with others, such as our neighbours, or members of the same community, or fellow citizens, or followers of the same religion. Our focus on particular identities can enrich our bonds and make us do things for each other and can help us to take us beyond our self-centred lives … [But] [t]he well-integrated community in which residents do absolutely wonderful

things for each other with great immediacy and solidarity can be the very same community in which bricks are thrown through the windows of immigrants who move into the region from elsewhere. The adversity of exclusion can be made to go hand in hand with the gifts of inclusion.55

Sen's last sentence is telling – 'can be made to go hand in hand'. Not must, not inevitably. Sen's invocation of diverse difference (within individuals, as within nations) and his ringing defence of the freedom to think critically about our multiple identities does not do the hard work – the line-drawing and the rule-making – that constitutional law requires.

Here, at least, is one place where the Constitutional Court's jurisprudence offers us useful guidance as to how the state ought to engage the religious, cultural and linguistic communities that make up the state and how those communities ought to engage one another. In Fourie, the Constitutional Court found that while the state could not continue to enforce common law rules and statutory provisions that prevented same-sex life partners from entering civilly-sanctioned marriages, the Constitution of the Republic of South Africa, 1996 had nothing immediate to say about religious prohibitions on gay and lesbian marriage and could not be read to require religious officials to consecrate a marriage between members of a same-sex life partnership.56 The Fourie court wrote:

[The amici's] arguments raise important issues concerning the relationship foreshadowed by the Constitution between the sacred and the secular. They underline the fact that in the open and democratic society contemplated by the Constitution … the religious beliefs held by the great majority of South Africans must be taken seriously … For many believers, their relationship with God or creation is central to all their activities. It concerns their capacity to relate in an intensely meaningful fashion to their sense of themselves, their community and their universe. For millions in all walks of life, religion provides support and nurture and a framework for individual and social stability and growth. Religious belief has the capacity to awaken concepts of self-worth and human dignity that form the cornerstone of human rights. Such belief affects the believer's view of society and founds a distinction between right and wrong … For believers, then, what is at stake is not merely a question of convenience or comfort, but an intensely held sense about what constitutes the good and proper life and their place in creation … Religious bodies play a large and important part in public life, through schools, hospitals and poverty relief programmes. They command ethical behaviour from their members and bear witness to the exercise of power by state and private agencies; they promote music, art and theatre; they provide halls for community activities, and conduct a great variety of social activities for their members and the general public … In the open and democratic society

55 Sen (n 54 above) 2–3. For a substantially more optimistic view about our capacity to sustain political institutions, in heterogeneous societies, through rational discourse, see KA Appiah Cosmopolitanism: Ethics in a world of strangers (2006) 113 144.
56 Minister of Home Affairs v Fourie (Doctors for Life International & Others, amici curiae); Lesbian & Gay Equality Project & Others v Minister of Home Affairs 2006 1 SA 524 (CC) paras 90-98.
contemplated by the Constitution there must be mutually respectful coexistence between the secular and the sacred.57

The *Fourie* court commits itself to five propositions that are fundamental for associational rights, generally, and for religious community rights, in particular. First, religious communities are a critical source of meaning for the majority of South Africans. Second, religious communities create institutions that support the material, intellectual, ethical and spiritual well-being of many South Africans. Third, religious associations, as part of civil society, play an essential role in mediating the relationship between the state and its citizens. Fourth, while religious, cultural and linguistic associations are entitled to articulate their 'intensely held world views', they may not do so in a manner that unfairly discriminates against other members of South African society. Fifth, although the 'intensely held world views' and practices of various religious communities by necessity exclude other members of South African society from some forms of membership and participation, such exclusion does not necessarily constitute unfair discrimination. Indeed, the *Fourie* court's decision makes it patently clear that, to the extent that exclusionary practices are designed to further the legitimate constitutional ends of religious associations, and do not have as their aim the denial of access to essential goods, the Constitution's express recognition of religious pluralism commits us to a range of practices that the Constitutional Court will deem fair discrimination. The refusal of some religious officials to consecrate same-sex life partnerships as marriages under religious law is but one form of fair discrimination. If we accept the *Fourie* court's fifth proposition, then South Africans are obliged to ask a number of other questions about the 'effects' of exclusionary practices. The easy question is whether communities possess the power to exclude members who initially agree to follow the rules of the community, but then subsequently refuse to do so. Of course, they do. The hard question comes in two related forms: (a) Would South African society be better off if it eliminated those exclusionary practices that removed 'non-compliant' individuals from the religious community; and (b) Would South African society be better off if it prevented other individuals – who begin as outsiders – from ever gaining entrance into the religious community they wish to join?

The answer to these hard questions turns primarily on access to those goods which individuals require in order to flourish. In twenty-first century democratic constitutional states, there are no hard and fast lines between the public sphere and the private sphere (and the various goods they provide to individuals.) The Constitution affords no easy answers about what remains in the private domain and thus subject to some constitutional pre-commitment to non-interference. Instead, the Constitution is primarily

57 As above.
Chapter 4

concerned with questions about individual and group access to the kind of goods that enable us to lead lives worth living.

Thus, the hard question for the immediate purpose of this article is this: When does religious discrimination constitute a justifiable impairment of the dignity of some of our fellow South Africans? That, I argued at the outset, was the question that we needed to answer: The hypothetical scenario adumbrated in the introduction offers a better opportunity – than Strydom does – to construct a cogent reply.

Before we answer the 'hard question', let us briefly rehearse the philosophical foundations for the kind of 'constitutive liberalism' that I have built up over the preceding pages. We began with the simple proposition that 'meaning makes us' – as individuals and as communities our social endowments largely determine who we are and what gives our lives meaning. In turning to concerns about 'capture', we recognised that freedom of religion, freedom of association and the right to communal religious practice are meant to protect the various well-springs of meaning that 'make us'. They do so by granting religious associations – amongst other associations – the right to determine the rules of membership, voice, exit that govern a community. At the same time, we recognised that associations of various kinds – religious associations included – enable us to create bridging and bonding networks that generate 'social capital'. This social capital – predicated as it is on trust, respect and loyalty – allows us to (a) maintain the practices and institutions that give our lives meaning; (b) create new practices and institutions that lead, potentially, for even greater flourishing; and (c) use these stores of real and figurative capital to transform 'exclusive' practices and institutions into non-exclusive practices and institutions that enable 'new' and perhaps 'historically disadvantaged members' of our society to benefit from the various goods these institutions distribute.

Given the above precepts of 'constitutive liberalism', it is easy to conclude that golf clubs that have been the bastion of white male Christian privilege must open their doors to persons of all colours, all sexes and all religions. But what of religious secondary schools that discriminate on the basis of an applicant's willingness to accept a prescribed religious curriculum and, at the same time, offer a better education than that generally available in our public schools? It would be foolish to dismantle such institutions solely on the grounds that either some form of exclusion takes place or that some re-inscription of privilege occurs. Almost all meaningful human labour occurs within the context of self-perpetuating social networks of various kinds. Taking a sledgehammer to social institutions that create and maintain large stores of real and figurative capital is a recipe for a very impoverished polity. The hard question thus turns on the extent to which religious communities can engage in justifiable forms of discrimination in the furtherance of constitutionally legitimate ends and the extent to which the state and other social actors can
make equally legitimate claims on the kinds of goods made available in these communal formations that cannot be accessed elsewhere.

Now the 'hard' question: Can the Pope legitimately discharge a Catholic organist from a Catholic church on the grounds that her sexual orientation as practised is inconsistent with the tenets of the faith even though she (unlike Strydom) desires religious inclusion and can make out a prima facie case that her dignity has been impaired? The answer must be 'yes'.

Why? The Catholic Church is the literal and figurative home to a billion people across the planet and millions across South Africa. It provides the setting for which so much of the meaningful action in the lives of its parishioners takes place. Moreover, the bonding networks and the bridging networks created by the church – for example, schools, churches, hospitals, charities, universities, non-governmental organisations – enrich the communities of believers and, more importantly, non-believers in which they are situated. Finally, it is one thing to allow the members of the church to carry out debates regarding the basic tenets of Catholicism. It is quite another to allow a state to smash open the doors of sanctuary and dictate – to the Pope – who may belong to the church and what rules they must follow. However outré one might believe the tenets of Catholicism to be, one cannot claim to be a defender of liberal 'constitutive' constitutionalism and simultaneously claim that the church must alter its belief set and practices so that they are consistent with the secular belief and practice sets of atheists like me who find the church's stances on reproductive rights, sexual orientation and stem cell research morally repugnant. Yes, a cost attaches to defending such a liberal constitutional order. However, if we wish to maintain meaningful settings for social action and, more importantly, leverage the social capital to be found in various associations and religious communities (by opening them up without destroying them), then we will have to tolerate beliefs and practices anathema to many of us. So long as church rules clearly preclude openly gay and lesbian members of the church from participating in public practices in the church, neither our courts nor our state has any business using such blunt cudgels as the right to dignity to reinstate an openly lesbian organist so that she might teach in a faith-based school and be remunerated for her efforts.
CHAPTER 5

MIGRATION, STREET DEMOCRACY AND EXPATRIATE VOTING RIGHTS*

Wessel le Roux**

1 Introduction

The nature and benefits of democratic citizenship are undergoing fundamental changes as the old Westphalian division of the world into independent nation states is slowly giving way to new forms of empire, transnationalism, postnationalism and cosmopolitanism. As a result of political instability, ecological change and economic globalisation, states are increasingly forced to reconfigure their relationships with migrant populations that are growing both in size and in political mobilisation. These migrant populations frequently include a combination of economic migrants and refugees. From the perspective of the nation state, these liminal figures have always constituted temporary exceptions that must, at one point or another, be restored to a position of normality.

Normality in this context means the restoration of the full range of rights accompanying nation state citizenship, either by voluntarily returning, or being returned (repatriated), to the state of origin, or, by being fully assimilated and integrated into the host state. The local integration of migrants is often described as a triangular process with legal, economic and social dimensions.1 As a legal process, local integration means that

* Previous versions of this essay were read as papers at a number of different seminars and conferences. The first version was presented on 5 August 2009 at the States of Statelessness Conference, Unisa. Reworked versions were later presented on 23 September 2009 at the SAIFAC seminar series, Is this seat taken?, and on 29 September 2009 at the Forced Migrations Studies Project (FMSP), Wits. I wish to thank Ulrike Kistner, Stu Woolman and Tamlyn Monson who initiated or organised these events. A special word of thanks to Karin van Marle for her support and insightful comments on the papers that were presented at the first and second events mentioned above. This paper was previously published in (2009) 24 SAPL 370.

** Professor, Faculty of Law, University of the Western Cape.

migrants are granted a progressively wider range of civil, social and political rights and entitlements by the receiving state. According to Crisp, the test for successful legal integration, especially in the case of refugees, is whether migrants have acquired permanent residence status and, ultimately, through naturalisation, full citizenship and voting rights in their host country.²

Against this standard model of local integration, it is important to insist (with Hannah Arendt and two of her most perceptive readers, Jean-François Lyotard and Giorgio Agamben) that the process of legal integration should not simply be equated with naturalisation or reduced to the gradual acquisition of traditional nation state citizenship.³ The idea of legal integration should rather be theorised as a disruptive and critical notion of disintegration, involving initially what Benhabib calls the gradual ‘disaggregation of citizenship’.⁴ Benhabib uses this phrase to describe the process whereby the cluster of rights that was traditionally associated with nation state citizenship is gradually broken apart and then distributed right by right to non-citizen migrants as well.

The disaggregation of citizenship is already reflected in the South African Bill of Rights in the fundamental distinction between constitutional rights that attach to ‘every person’ and constitutional rights that attach only to ‘citizens’. This distinction has given rise to a number of important Constitutional Court cases. Most fundamentally, in Lawyers for Human Rights v Minister of Home Affairs, the Court rejected the claim by the government that arriving refugees, who were not yet legally admitted into

² Crisp (n 1 above) 1.
³ H Arendt The Jew as Pariah: Jewish identity and politics in the modern age (1978) 55 famously criticised the desire of Jewish refugees during World War II to present themselves to their host countries as perfectly assimilated, prospective and patriotic citizens. Arendt found the desire to take refuge in nation state citizenship disconcerting, because it involved a lack of political courage to fight for a change in the social and legal status of stateless Jews in Europe. She sought to counter the problematic Jewish identity of the newly assimilated social parvenus by embracing the Jewish counter-tradition of conscious social pariahs or outlaws (represented by Rahel Varnhagen, Franz Kafka and Charlie Chaplin). In the process, Arendt famously remarked that these conscious outlaws gained the world of politics and represented the new political avant-garde of Europe. J-F Lyotard Toward the postmodern (1993) 144 welcomes the fact that here Arendt links politics to a radical critique of representation (the Jewish God as Word that de-installs) which leaves the subject in a political state of ‘being-together without roots’, a togetherness that behaves according to the formal demands of (aesthetic) judgment alone. G Agamben Means without end: Notes on politics (2000) 14 equally celebrates Arendt’s critique of the paradigm of nation state citizenship as alternative to the statelessness of Jewish refugees. Agamben, with Arendt, that the stateless refugee represents the central figure of our political history. He actively calls for a process by which each citizen recognises the ‘refugee that he or she is’ and gradually turns him or herself into a denizen of cities that perforate and deform the space of the nation state. I return later in this essay to this reinterpretation of statelessness, not as a temporal abnormality, but as the precondition for the political survival of humankind.
the country, could not be beneficiaries of the rights in the Constitution.\(^5\) In *Khosa v Minister of Social Development*, the Court held that the constitutional right to social security also applies to citizens of Mozambique who have permanent resident status in South Africa.\(^6\) Finally, in *Union of Refugee Women v Director, Private Security Industry Regulatory Authority*, a minority of the Court held that the reservation of work in the security industry to citizens and permanent residents violated the constitutional rights of refugees.\(^7\) Given this jurisprudence, the crucial question that follows next is whether the disaggregation of post-apartheid citizenship should be extended beyond civil and social rights to political rights as well. Benhabib believes that it should. According to her, the disaggregation of citizenship reveals its real critical potential precisely where claims are made for an extension of the process to include voting rights as well.\(^8\) At this point, the legal and political integration of migrant populations fundamentally disrupts our traditional understanding of nation state citizenship as a form of political membership and action.

In this essay I begin to investigate the influence of migration on citizenship and the allocation of voting rights from the alternative model of legal (dis)integration suggested by Benhabib. I do so by presenting a detailed analysis of two recent judgments in which the issue of migrant voting rights was served before the South African Constitutional Court for the first time.\(^9\) The discussion most immediately involves the question of whether resident citizens retain the constitutional right to vote in general elections. I argue with the Constitutional Court that they do. The discussion inevitably includes the question of whether non-resident citizens (expatriates) retain the constitutional right to vote in the general elections of the Republic. I argue (some would say against the Constitutional Court) that they do not.

---

\(^5\) [2004] ZACC 12; 2004 4 SA 125 (CC); 2004 7 BCLR 775 (CC).
\(^6\) [2004] ZACC 11; 2004 6 SA 505 (CC); 2004 6 BCLR 569 (CC).
\(^8\) According to Benhabib, the disaggregation of citizenship rights is a political process that is centred on the democratic iteration of cosmopolitan norms; a process that finds its clearest expression in the debates around migrant citizenship and voting rights in Germany and the European Union. As an example of democratic iteration, Benhabib discusses the enactment of a new citizenship law by the German Federal Parliament in January 2000, in direct response to the 1990 judgment of the German Constitutional Court in which a law that granted resident non-citizens the right to vote in municipal elections was declared unconstitutional (Benhabib (n 4 above) 62 - 67; BVerfG 83, 37 Ausländerwahlrecht Schleswig-Holstein). Even before the adoption of the new citizenship law, the judgment of the Court had been overtaken by the Maastricht Treaty of 1993, which granted the right to vote in municipal elections to all European citizens on the basis of residence alone. Benhabib (n 4 above) 35 goes as far as claiming that critical theory finds its contemporary expression in the drive for the extension of voting rights as part of a new ‘politics of cosmopolitan membership’. This politics concentrates on rights litigation that burdens courts with the task of ‘negotiating the complex relationship between rights of full membership, democratic voice and territorial residence’.
\(^9\) Richter v Minister for Home Affairs [2009] ZACC 3; 2009 5 BCLR 448 (CC); AParty v Minister for Home Affairs; Moloko v Minister for Home Affairs [2009] ZACC 4; 2009 6 BCLR 611 (CC).
Once the distinction between resident and non-resident citizens is accepted as constitutionally mandated by the value of democratic accountability, it becomes possible to radicalise the idea of resident nation state citizenship into a form of post-national and nationalist denizenship. Throughout the discussion my real concern therefore lies with the effect that the construction of emigrant citizenship and the allocation of absentee voting rights might have on the continued disaggregation of citizenship as a critical process, or the allocation of voting rights to various classes of resident non-citizens. I assume, without arguing the case here, that the same constitutional principle that governs the allocation of voting rights to non-resident citizens at one end of the spectrum should also govern the allocation of voting rights to noncitizen residents at the other. The key question is whether that principle is to be located in the idea of global citizenship and diasporic nationalism (as the Constitutional Court seems at times to suggest), or rather in a robust conception of local or street democracy.

The discussion is framed by two important recent events. On 16 August 2005, the Protocol on the Facilitation of Movement of Persons was adopted by the Southern African Development Community (SADC). The Protocol explicitly recognises the right to migrate and to economically establish oneself throughout the SADC region. However, the Protocol does not link these newly acquired migration rights to any post-national notion of community citizenship. It is equally silent about the political status and rights of economic migrants, both in their host and sending countries. In more ways than one, the Protocol therefore echoes the limits of the African Common Position on Migration and

---

10 I do not question here whether voting retains any link to transformative politics and whether the focus on voting rights inevitably inscribes the argument within a national constitutional paradigm. Ample material exists for such an argument. M Hardt & A Negri Empire (2000) 210, eg, locate the future of republican politics in the anti-institutional ‘will to be against’ that characterises the ‘multitude’ as opposed to ‘the people’ of classical republican thought. Compare in this regard the idea of ‘constitutional mob action’ as a legitimate form of political participation celebrated by LD Kramer The people themselves: Popular constitutionalism and judicial review (2004) 25 - 27. The importance of the right to vote in transformative politics surfaced in a less radical form in the recent jurisprudence of the Constitutional Court. One of the reasons presented by Yacoob J in Doctors for Life International v Speaker of the National Assembly [2006] ZACC 11; 2006 12 BCLR 1399 (CC); 2006 6 SA 416 (CC) for refusing to read a participatory element into the law-making process, was precisely that it would diminish the value of the right to vote that is said to form the focus of liberation and post-liberation politics.


Migration recently adopted by the African Union.\textsuperscript{13} Like the Protocol, the Common Position both embraces economic migration as a development strategy and restricts itself to the economic rights of Africa’s diasporic communities.

The danger that this one-sided economic or development approach might pose to the well-being of migrant populations was sadly underlined by the wave of xenophobic violence that broke out in South Africa on 11 May 2008. In two long weeks the violence escalated to such an extent that the military had to be deployed in residential areas to curb the civil unrest. By the time it was eventually brought under control, the violence had left more than 60 persons dead, 670 injured and 100 000 displaced. According to one study of the events, the outbreak of the xenophobic violence can be attributed to a breakdown of democratic governance and participatory politics at local government level.\textsuperscript{14} As such, the violence underscores the urgent need to explore ways to legally and politically integrate migrants through an inclusive notion of post-national and, even more importantly, a post-nationalist notion of local democratic citizenship, or more specifically, street democratic denizenship.

2 \hspace{1em} Formulating the basic question

Any suggestion that a qualitative distinction can be drawn between resident citizenship and non-resident citizenship for the purpose of the allocation of voting rights, must from the outset deal with the fact that the Constitution does not itself distinguish between resident and non-resident citizens. It simply states that all citizens are ‘equally entitled to the rights, privileges and benefits of citizenship’.\textsuperscript{15} These benefits include ‘the right to vote in elections for any legislative body established in terms of the Constitution’.\textsuperscript{16} In this manner the Constitution gives comprehensive recognition to the founding constitutional value of ‘universal adult

\begin{footnotesize}
\begin{enumerate}
\item ['African Common Position on Migration and Development’ adopted by the Executive Council at its 9th ordinary session, 25 - 29 June 2006, Banjul, The Gambia (EX CL/227 (IX)).
\item Sec 19(3) of the Constitution. Unequivocal as these provisions might be, they do not comprehensively define who citizens are. The Constitution explicitly leaves it to parliament to determine who are citizens and merely imposes a duty on parliament to adopt national legislation to ‘provide for the acquisition, loss and restoration of citizenship’ (sec 3(3)). See J Klaaren ‘Citizenship’ in Woolman et al Constitutional law of South Africa (2008) 60-1.
\item Sec 19(3) of the Constitution. Unequivocal as these provisions might be, they do not comprehensively define who citizens are. The Constitution explicitly leaves it to parliament to determine who are citizens and merely imposes a duty on parliament to adopt national legislation to ‘provide for the acquisition, loss and restoration of citizenship’ (sec 3(3)). See Klaaren (n 15 above) 60-1.
\end{enumerate}
\end{footnotesize}
Chapter 5

suffrage'.17 In the ACDP case, the Constitutional Court accordingly laid down the general principle that when interpreting provisions in electoral statutes, courts must ‘seek to promote enfranchisement rather than disenfranchisement and participation rather than exclusion’.18

It was precisely on the basis of these inclusive provisions and pronouncements that Willem Richter launched his urgent application in the North Gauteng High Court for an order to declare section 33(1)(e) of the Electoral Act 73 of 1998, which regulates the granting of absentee votes, unconstitutional.19 Mr Richter is a 27 year-old South African citizen who grew up in Pretoria, qualified as a teacher at the University of the North-West, served in the Democratic Republic of the Congo as a member of a South African peace-keeping force, but currently lives and works in England as a teacher. Towards the end of 2008, Mr Richter approached the South African Electoral Commission to find out whether and how he could vote in the (then) upcoming 2009 general elections. He was informed that the only way for him to vote was to return to South Africa and to vote in the voting district where he was last registered as a voter. This was so because section 33(1)(e) of the Electoral Act restricted absentee voting rights to holiday makers, businessmen, students and sportsmen, who were temporarily outside the country.20 Mr Richter's temporary absence from the Republic, due to his employment in England, thus fell outside the limited scope of the section.

The predicament of Mr Richter and many similarly-positioned citizens was soon taken up by a number of opposition parties and civil society organisations. On 22 January 2009, former President FW de Klerk requested the Commission, in the name of the Centre for Constitutional Rights, to re-interpret the provisions of section 33(1)(e) in a manner that would allow all citizens living and working abroad to vote in the national elections.21 When this request was again formally turned down, the

---

17 Sec 1(d) of the Constitution.


20 The section reads as follows: ‘The Commission must allow a person to apply for a special vote if that person cannot vote at a voting station in the voting district in which the person is registered as a voter, due to that person's temporary absence from the Republic for purposes of a holiday, a business trip, attendance of a tertiary institution or an educational visit or participation in an international sports event, if the person notifies the Commission within 15 days after the proclamation of the date of the election, of his or her intended absence from the Republic, his or her intention to vote, and the place where he or she will cast his or her vote’ (my emphasis). During the subsequent litigation, the Minister of Home Affairs, interestingly enough, disputed whether the narrow interpretation of sec 33(1)(e) by the Electoral Commission was correct. The Minister argued against the Electoral Commission that sec 39(2) of the Constitution requires a broad interpretation of the term 'business trip' to include any temporary employment overseas. The Constitutional Court rejected this interpretation and confirmed the interpretation of the Act by the Electoral Commission. I do not explore this issue any further in this note. See L Nkuna ‘The right to vote, section 39(2) and the limits of conforming interpretation under the Electoral Act 73 of 1998’ (2009) 24 SAPL 638.

21
Freedom Front Plus announced that it would assist Mr Richter to challenge the constitutionality of section 33(1)(e) in court. The High Court (per Ebersohn AJ) eventually granted the application brought by Mr Richter with costs. In the process, the Court ordered the severance of both the limited number of voter activities listed in the section (call this the first severance order) and the requirement that only voters who are temporarily absent from the Republic may apply for special absentee votes (call this the second severance order).

The effect of the first severance order was that Mr Richter could vote in the upcoming 2009 general elections at a voting station in London. Temporary absence from the Republic, for whatever reason, no longer adversely affected the voting rights of citizens who were already registered as voters. The effect of the second severance order was even more far-reaching. By scrapping the requirement that only registered voters who were temporarily outside the country could vote overseas, the Court opened the door for citizens who had permanently left the country (expatriates) to also vote overseas in the general elections. It was this secondary effect of the judgment, rather than Mr Richter's far narrower predicament, which captured the headlines and websites in South Africa. Regret was immediately expressed that Mr Richter had not also challenged the residence based registration requirements of the electoral system as a whole, so that the voting rights of non-resident citizens could be further extended to expatriate citizens who were not registered as voters.

As it turned out, the perceived limitation of the Richter litigation was of little import, for when the Richter case was served before the Constitutional Court for confirmation a few weeks later, the dynamics of the case had completely changed. A number of political parties and interest groups had applied and were granted leave to join the confirmation proceedings, either as intervening parties or friends of the court. A couple of new applicants also made use of the opportunity to approach the Court for direct access on issues relating to absentee voting rights. The newly-added parties to the dispute had a far broader political agenda than Mr Richter. They explicitly sought an order from the Court to affirm the constitutional right of non-resident or expatriate citizens to register as voters.

---

22 In the end, more than 16 000 South African citizens voted overseas on 15 April 2009.
24 N de Haviland ‘The right of South African’s [sic] living abroad to vote’ www.cfcfr.org.za/?p=416 (accessed 11 May 2009). De Haviland is the Director of the Centre for Constitutional Rights of the FW de Klerk Foundation. The ‘ordinary resident’ requirement underlying the Electoral Act, and the electoral system as a whole, makes the registration of non-resident or expatriate citizens as voters impossible. Sec 8(3) of the Act provides that a person’s name ‘must be entered in the voters’ roll only for the voting district in which that person is ordinarily resident and for no other voting district’. In this regard, sec 7(3)(a) of the Act defines ‘ordinary residence’ as the ‘home or place where that person normally lives and to which that person regularly returns after any period of temporary absence’.
voters for, and then vote in, South African national and provincial elections. What the newly-added parties to the Richter litigation wanted was nothing less than for the Court to rule that the ordinary residence requirement underlying the South African electoral system was unconstitutional.

In response to this broadened and politicised agenda, the Minister of Home Affairs and the Electoral Commission both urged the Court not to rule on the constitutionality of the residence requirement on an urgent basis and as a court of first and last instance. The Commission also made use of the opportunity to remind the Court that the design of the electoral system was constitutionally left to the discretion of the legislature, and that the ordinary residence requirement formed the backbone of the very same system of proportional representation that the Constitution itself prescribes. The Electoral Commission added that the temporary absence requirement for absentee voters in section 33(1)(e) of the Electoral Act was merely the flipside of the ordinary residence requirement for voter registration in section 8(3) of the Act. The Electoral Commission insisted that, if the Court could not at this stage rule on the constitutionality of the residence requirement in section 8(3), then it could also not at this stage rule on the constitutionality of the residence requirement in section 33(1)(e). Thus, the Constitutional Court could, at best for the various applicants, confirm the first severance order of the High Court.

The Constitutional Court agreed with much of the Minister’s and the Commission’s argument. It specifically held that the constitutionality of the residence requirement underlying the electoral system could not be decided at this early (or late) stage, depending on how one looked at the matter. The Court also accepted that the design of the electoral system was ‘a matter that lies peculiarly with Parliament’s constitutional remit’. In spite of these agreements, the Court nevertheless proceeded to confirm the second severance order of the High Court by striking the words ‘temporary absence’ from section 33(1)(e) of the Act.

The basic question that I am grappling with is how the Constitutional Court could possibly have done what the Minister and Electoral

---

25 These parties also challenged the 15 days notice period in sec 33(1)(e) of the Electoral Act. The Court dismissed this application. I do not discuss this aspect of the judgment in any further detail. See Richter (n 9 above) para 80 - 84.

26 AParty (n 9 above) paras 21 - 22.

27 Second Respondent’s Heads of Argument para 74.

28 AParty (n 9 above) para 68.

29 AParty (n 9 above) para 71.

30 AParty (n 9 above) paras 72 & 80. The Court held that it was too late for it to be approached on an urgent basis so shortly before the elections; the Court also held that it was too early in the public debate and litigation process for it to be directly accessed on the matter.

31 AParty (n 9 above) para 80.

32 AParty (n 9 above) paras 33 - 34.
Commission thought was impossible, namely, to rule on a principled basis on the constitutionality of the temporary absence requirement in section 33(1)(e) of the Electoral Act, while declining to rule on the constitutionality of the permanent residence requirement in section 8(3) of the Electoral Act. Why did the Court refuse to accept the constitutional validity of a distinction between resident and non-resident citizenship for the sake of section 33(1)(e)? Should this refusal be understood as an insurmountable constitutional bar to the development of a post-apartheid theory of democratic denizenship? The search for possible answers to these questions must begin with a closer reading of the initial High Court judgment.

3    Richter in the High Court

In his original notice of motion in the High Court, Mr Richter applied for an order declaring that all registered voters who were working and living outside the country, whether temporarily or not, have the right to cast a special absentee vote in South African general elections. The broad scope of the application rested essentially on an equality argument, directed at the unequal treatment between various categories of registered voters. Mr Richter’s main concern in this regard was not with the differentiation within sub-section 33(1)(e) between different classes of temporarily absent voters (students and temporary employees, for example), but the differentiation within section 33(1) as a whole between different classes of absentee voters. The section as a whole created two classes of absentee voters. Section 33(1)(b) dealt with the first class and granted the right to cast a special vote to voters who could not vote in their voting district due to ‘absence from the Republic on government service or membership of the household of the person so being absent’. As noted above, section 33(1)(e) dealt with the second class and granted a right to cast a special vote to voters who could not vote in their voting district because of a ‘temporary absence from the Republic for purposes of a holiday, a business trip, attendance of a tertiary institution or an educational visit or participation in an international sports event’.

Mr Richter pointed out that section 33(1)(b) of the Electoral Act allowed people who work overseas in government service to cast special votes abroad, while section 33(1)(e) denied the same right to people who work overseas in the private sector. Alerted to this fact, the High Court ruled that this differentiation was an unacceptable form of discrimination in that it affords a privilege to people on government services as opposed to people in the private sector and amounts to totalitarianism on the part of the state.33

33  AParty (n 9 above) para 61.
From this starting point, a number of conclusions quickly followed. The first was that the state could not justify the unfair discrimination against temporarily absent voters in private employment with an appeal to logistical or administrative concerns. Provision for special voting was already made for people working in government service at diplomatic missions across the world. There was no reason why the existing arrangements could not simply be extended to include other classes of absentee voters as well. In fact, given the weakness of any logistical justification for the discrimination, the Court all but agreed with Mr Richter and the expatriate voting rights lobby that the disenfranchisement of voters who live and work overseas formed part of a vindictive policy of retaliation against the unpopular emigrant minority:

It must be noted that in terms of the then current legislation namely section 25 of the Electoral Act, 202 of 1993, South African citizens who were outside the country during the 1994 elections could vote at foreign voting stations without even having to apply for special votes. No distinction was drawn between different classes of citizens (ie government employees or privately employed citizens) [...] The effect of the [subsequent legislative changes], watered down the right to vote which is inexplicable, arbitrary and inconsistent with the word and spirit of the fundamental right to vote contained in the Constitution.

The Minister of Home Affairs was, herself, partly to blame for the starkness of the conclusion reached by the Court. In none of the Richter cases did the Minister try to defend a residence based approach to absentee voting rights as a matter of democratic principle or policy. In fact, the Minister completely failed to file any heads of argument in the High Court application (in spite of repeated requests by the Court to do so). In the Constitutional Court, where the Minister finally filed heads of argument, the only democratic principle that the Minister could offer was a separation of powers argument, namely, that the design of the electoral system fell squarely within the jurisdiction of the democratic legislature. In the absence of a convincing argument based on the constitutional value of democracy itself, Ebersohn AJ proceeded as if the only possible

---

34 *AParty* (n 9 above) para 77.
35 *AParty* (n 9 above) para 12 read with paras 45 - 48. Ebersohn AJ believed that the same conclusion was underscored by the genesis of the provision. In this regard, he relied extensively on a remark made by the then Minister of Home Affairs, Mangosuthu Buthelezi of the Inkatha Freedom Party, during the Second Reading Debate of the Amendment Bill: ‘When I spoke to this House during my Department’s budget debate, I had committed myself to make provision to enable South Africans abroad to vote. Accordingly, in the Electoral Law Amendment Bill which I brought to Cabinet provision was made for all South Africans abroad to vote, irrespective of their reasons to be abroad, however, Cabinet chose not to adopt my proposal [...] I felt that as a matter of policy it was essential to allow all South Africans an opportunity to participate in the electoral process. However, what in my opinion was not constitutionally problematic, may become so as the Bill before us now only allows certain citizens who are abroad for certain temporary purposes to vote, while depriving others who are in a similar situation of the same opportunity.’
36 *AParty* (n 9 above) paras 7 28.
justification for the various restrictions on absentee voting rights had to be logistical or administrative in nature.

This assumption should not blind us to the fact that the ill-conceived section 33(1)(e) remained embedded within a larger residence based approach to the electoral system, as reflected in section 8(3) of the Electoral Act. Section 33(1)(e) set out to combine three elements of the residence based electoral system, but ended up being little more than a legislative self-contradiction. The section provided that only registered voters (the first element: registration) who were temporarily absent (the second element: residence) for a selected number of reasons (the third element: reason) could apply for special absentee votes. The obvious weakness was the third element. It had no principled connection to the larger residence based electoral system, but formed part of a misguided attempt to use the Electoral Act to counter emigration from the Republic. Once the Court had accepted that the equality provisions of the Constitution required that diplomats (section 33(1)(b)) and other absentee voters (section 33(1)(e)) should be placed on exactly the same footing, the second element (temporary absence or residence) also became problematical. As this requirement did not apply to absentee voters in the government service category of section 33(1)(b), Ebersohn J held that it could also not apply to the absentee voters in the section 33(1)(e) category. What was more, because there was no further reason to distinguish between the two different classes of absentee voters, the High Court ordered that section 33(1)(b) be deleted from the Electoral Act and that section 33(1)(e) be extended to cover all registered voters who are absent from the Republic on election day (including those in government service). It was in order to secure this extension, and for no other reason, that the Court ordered the deletion of the temporary absence requirement (the second severance order) in addition to the deletion of the listed categories (the first severance order). All absentee voters, diplomats and expatriates alike, would in future equally qualify in terms of the same statutory provision for a special absentee vote in national elections.

It is not necessary to explore the weakness of the High Court’s equality argument here. Nothing in my discussion turns on it. Suffice to say that

37 *AParty* (n 9 above) para 81.
38 This equalisation meant that voters in the former diplomatic service category were effectively disenfranchised as far as provincial government elections were concerned. Before the High Court judgment, absentee votes in terms of sec 33(1)(b), read with reg 9 of the Election Regulations (GN R12 in GG 25894 of 2004-01-07), could be cast in both national and provincial elections (on the strength of a deeming provision that all such votes were cast in Pretoria and thus in the Gauteng Province). By contrast, absentee votes in terms of sec 33(1)(e), read with reg 12, were restricted to national government elections only. The High Court did not deal with this effect of its order. The Constitutional Court reinstated the right of absentee voters in government service to vote in both national and provincial elections (para 93). However, it refused to rule on the right of other absentee voters to vote in provincial elections (paras 85 - 92). I return to this point later in the discussion.
because of the obvious difficulties with the argument, the Constitutional Court deliberately decided not to endorse the High Court’s equality-based approach to the right to vote. Given the rejection of the equality argument, the Constitutional Court understandably also refused to confirm the deletion of the section 33(1)(b) diplomatic service category from the Electoral Act. This refusal, although seemingly only an incidental feature of the Court’s judgment, is more significant than it might appear at first sight. Recall that in Ebersohn AJ’s mind, the deletion of section 33(1)(b) and the severance of the temporary absence requirement from section 33(1)(e) formed two sides of the same equality coin. By refusing to confirm the first deletion, the Constitutional Court effectively removed the whole basis of Ebersohn AJ’s claim that the temporary absence requirement should also be deleted.

This returns us to the basic question posed above: On what basis was it thereafter possible for the Constitutional Court to proceed to rule that the second severance order of the High Court was both necessary and correct? On what constitutional basis did the Court reject the statutory distinction between resident and non-resident citizens for the purposes of absentee voting rights?

4  Richter in the Constitutional Court

As a starting point for further discussion, it is perhaps best to quote the complete judgment of O’Regan J on behalf of the Constitutional Court as far as it pertains to the second severance order and the voting rights of non-resident citizens:

The High Court also severed the word ‘temporary’ from the section. In my view, this order is correct.41

Incredible as it may sound, this is the full, reasoned judgment of the Constitutional Court on the point. This single statement appears towards the end of the judgment, just after O’Regan J had confirmed the first severance order of the High Court and just before she restated the new wording of section 33(1)(e) after the two severances had been effected. It is therefore not immediately clear from the judgment precisely what the constitutional defect was that had to be remedied by the second, as opposed to the first, severance order. The Court had already rejected the equality argument of the High Court, on the one hand, and the applications for direct access on the constitutionality of the residence based registration system, on the other (in the AParty case). The problem could not be, therefore, that the temporary absence requirement incorporated an

---

39  Richter (n 9 above) paras 78 & 99.
40  Richter (n 9 above) para 93.
41  As above.
unconstitutional residence based restriction on the right to vote (the *AParty* argument), or that it did not equally apply to all absentee voters (the Ebersohn AJ argument). If there was a constitutional problem with the temporary absence requirement in section 33(1)(e), it had to lie elsewhere. The *Richter* (and *AParty*) judgments suggest four possible reasons why the distinction between resident and non-resident citizens in section 33(1)(e) of the Electoral Act might be unconstitutional.

4.1 The right to vote and section 33(1)(e) of the Electoral Act

The first line of reasoning is O’Regan J’s analysis of section 33(1)(e) itself. O’Regan J dealt between paragraphs 60 and 70 of her judgment with the impact of section 33(1)(e) on the right to vote. She did so under the heading ‘the classes of absentee voters permitted a special vote’. She concluded in paragraph 70 that section 33(1)(e) limited the right to vote enshrined in section 19 of the Constitution ‘by restricting the classes of voters who are absent from the Republic on polling day from participating in elections’. On the face of it, both the heading to and conclusion of O’Regan J’s judgment refer to the four classes of voters that are specifically listed in the section (tourists, students, businessmen and sportsmen). However, there are at least two classifications or classes of absentee voters involved in section 33(1)(e). The first classification divides all the registered voters who are temporarily absent from the Republic according to the reasons for their absence. It is this classification that resulted in the list of arbitrarily selected voter activities that was rightfully removed from the section by the first severance order. The second classification divides registered voters into those who are temporarily absent from the Republic and those who are permanently absent from the Republic. It is this classification that was removed by the second severance order. The key question is thus whether this second classification was included under the ‘classes of voters’ that O’Regan J considered during her rights analysis. Every indication is that it was not.

To begin with, O’Regan J explicitly referred to the Minister’s argument in the case as support for her conclusion that the limited list of voter activities in section 33(1)(e) violated the right to vote. However, what the Minister argued was that every registered absentee voter could be accommodated within the specific classes listed in section 33(1)(e) for ‘as long as he or she remains ordinarily resident in the country’ (my emphasis). O’Regan J was acutely aware that the Minister’s concession only pertained to the first of the classifications mentioned above. The Minister insisted throughout that the residence qualification is beyond reproach. There is no response by O’Regan J to the important qualification introduced by the Minister. In fact, O’Regan J seemed content to accept

---

42 *Richter* (n 9 above) para 61.
43 As above (my emphasis).
that the constitutional argument must be restricted only to those voter activities that were listed. This would explain why she continued along these lines by describing Mr Richter as a person who fell outside the listed classes mentioned in section 33(1)(e) but who ‘has not permanently left the country’ (my emphasis). By adding the highlighted qualification, O’Regan J again indicated that she was limiting herself to the first of the classifications mentioned above. In fact, her statement gives the impression that Mr Richter’s voting rights would indeed have been affected had he permanently left the country. O’Regan J dealt with the plight of Mr Tipper and Mr and Mrs Moloko, the other applicants in the case, in much the same manner. She again stressed that these registered voters considered themselves to be resident citizens of South Africa who were only temporarily outside the country.44 Once more the impression is created that this is indeed a decisive consideration in the case.

A second point to consider is that O’Regan J at times said in so many words that her sympathy towards the plight of all the applicants was closely tied to the fact that they were only temporarily absent from the country. She explained, for example, that their temporary absence will economically benefit the Republic ‘when they return’ with greater work experience or money to buy a house.45 When O’Regan J concluded her rights analysis with the statement that section 33(1)(e) limits the right to vote ‘by restricting the classes of voters who are absent from the Republic on polling day from participating in elections’, she could therefore only have been referring to the classes of voter activities listed in the section. The section limited the right to vote because it excluded some classes of registered voters who were temporarily absent from the country from casting special votes. The question of whether the section also limited the right to vote because it disqualified the class of expatriate or non-resident citizens did not receive any attention from O’Regan J during her rights analysis.

O’Regan J proceeded in the next eight paragraphs of her judgment to consider whether the limitation of the right to vote could be justified in terms of section 36 of the Constitution. The limitation analysis confirms that the disqualification of expatriate citizens was not the cause of her constitutional concerns in the matter. O’Regan J made this clear at the outset, by formulating the question involved as the possible limitation of the ‘under-inclusiveness of [...] the categories listed in section 33(1)(e)’.46 She then immediately indicated that there was nothing contained in the written argument of the Minister to justify ‘the restrictive classes contained in section 33(1)(e)’ and that the Minister conceded as much.47 O’Regan J correctly indicated that the main thrust of the Minister’s written argument

---

44 Richter (n 9 above) para 66.
45 Richter (n 9 above) para 69.
46 Richter (n 9 above) para 76.
47 Richter (n 9 above) para 72.
was directed against the registration of permanently absent or expatriate citizens and that this question did not arise in the Richter case. As O’Regan J pointed out herself, the crux of the Minister’s argument was that the distinction between resident citizens and expatriate citizens was constitutionally justifiable. It was on the basis of this distinction that the Minister argued that expatriate citizens did not qualify for special votes (under section 33(1)(e)) or for voter registration (under section 8(3)). As noted before, the only point that the Minister conceded was that all resident citizens or temporarily absent voters had a right to cast a special vote overseas. There was indeed nothing in the written argument of the Minister to justify any interpretation of section 33(1)(e) to the contrary. The same does not apply to the second classification mentioned above. The crucial point is that O’Regan J did not deem it necessary to deal with justification for the disqualification of the class of non-resident citizens that the Minister tried to provide.

The same applies to the argument by the Electoral Commission. Like the Minister, the Commission also objected to the second severance order while conceding the need for the first order. In as far as O’Regan J again relied on the fact that the Commission did not oppose the relief sought, she could only have been referring to the severance of the list of voter activities from the section. The comparative study undertaken by the Court during the limitations analysis was likewise explicitly directed at the tendency of democratic countries to allow voters to cast absentee votes without restrictions ‘on the basis of the activity undertaken abroad by the absent voters’ (my emphasis). When O’Regan J concluded in paragraph 78 that the limitation of the right to vote could not be saved by section 36 of the Constitution, this conclusion therefore related only to the classification of voters according to the activities that they were undertaking while overseas. Nothing in the limitation analysis touched on the question of whether the disqualification of the class of expatriate or non-resident citizens could be or even needed to be constitutionally justified.

When the Court proceeded to consider the question of an appropriate remedy for the constitutional defect in section 33(1)(e), it understandably confirmed the first severance order of the High Court and deleted the list of voter activities from the section. The deletion of this list was necessary to make the section universally applicable to all absentee voters who were temporarily absent from the Republic, regardless of the activity undertaken abroad by the absent voters. Given the narrow frame of the reasoning that preceded this severance order, this is where the judgment should have stopped. If O’Regan J’s analysis of section 33(1)(e) failed to reveal what constitutional defect the Court sought to remedy with the second

48 As above.
49 Richter (n 9 above) para 72.
50 Richter (n 9 above) para 75.
51 Richter (n 9 above) para 77.
severance order, perhaps her more general analysis of the nature and importance of the right to vote contains the ground for her judgment.52

4.2 The right to vote and section 19 of the Constitution

O’Regan J insisted that the right to vote must be approached from the perspective of dignity (as opposed to equality). Following the celebrated comments by Sachs J in the August case, O’Regan J described the right to vote as a ‘badge of dignity’ and a ‘symbol of citizenship’.53 From this dignity based perspective, the formal equality of political rights is not its greatest vice, as a long tradition of Marxists has claimed, but its greatest virtue. It provides the basic formula of instant nationhood and implies that all voting rights provisions must as far as possible be interpreted to favour the enfranchisement of as many people as possible.

It is certainly possible to push this dignity based approach to political rights to reach the conclusion that nationhood and citizenship formally and conceptually imply the right to vote as a necessary constituent.54 This is certainly how the expatriate lobby has embraced this portion of O’Regan J’s judgment.55 However, O’Regan J herself did not pursue this line of reasoning. She immediately qualified her dignity based approach by adding that the right to vote is not merely a symbolic icon of nationhood but an operative feature of a working multi-party democracy.56 The right to vote is thus inextricably linked to and internally limited by the right to free and fair elections.57 In as far as the Constitutional Court refused to entertain the argument that the residence based nature of the electoral system and registration process is unconstitutional, nothing that O’Regan J said or implied in her analysis of the right to vote can have any conclusive bearing on the merits of the second severance order.

We are seemingly no closer to answering the question of why O’Regan J declared the temporary absence requirement unconstitutional and issued the second severance order. Perhaps we might have more success if we step back from the case for a moment, and consider the interpretation that the Constitutional Court itself has placed on the judgment of O’Regan J.

4.3 The right to vote and the AParty case

In rejecting the applications for direct access in the AParty case, Ngcobo J had to deal with the argument that the issues in the AParty and Moloko
cases already served before the Court in the Richter case. Ngcobo J rejected this argument by carefully distinguishing the Richter case from the matter before him. He explained that the case before him dealt with the eligibility to register as a voter in South African law, while the Richter case dealt with the restrictions and burdens that could be placed on an already registered voter who wished to exercise his or her right to vote. By drawing this distinction between the right to register and the right to vote, Ngcobo J reveals that the Richter judgment rested on the principle that every registered voter must be given a reasonable opportunity to cast a vote in an election (which might mean the reasonable opportunity to cast a special vote overseas). Ngcobo J thus offers a third possible ground for the judgment in Richter. If we accept his interpretation, this would explain why O'Regan J never explicitly referred to ‘the class of registered voters who are ordinarily resident in the Republic’ but only spoke about the ‘class of registered voters’, or more descriptively, about ‘any citizen who wishes to vote in the election [and] is a registered voter’ as the target of her concern.

A rejection of the classification of registered voters into resident and non-resident citizens was thus implied in the Richter judgment from the start. If the class of all registered voters is taken as the frame of reference, as O'Regan J must now be assumed to have done, then the attempts to further classify voters according to their activities overseas or their residence status become constitutionally untenable. O'Regan J simply accepted as a self-evident constitutional principle that voter registration conclusively settled the right to cast a special vote. All that mattered thereafter was to ensure that all registered voters were given a reasonable opportunity to cast their vote. If the second severance order in the Richter case is understood in this way, it means that the judgment has nothing to do with the vindication of expatriate voting rights. The effect of the judgment might well have been that some expatriate citizens acquired the right to vote (which they did), but the reason they did so was purely because their names still appeared on the voters' roll, not because of any independent constitutional principle that citizenship includes the right of expatriate citizens to vote or to register as voters.

While this narrow reading of the second severance order might have some initial appeal (in as far as it at least provides us with a reason for the order), the fact of voter registration cannot by itself provide a reason for the second severance order. To begin with, registration as a voter is not a once-off event. There is a (widely neglected) civic and legislative duty on the

58 AParty (n 9 above) paras 52 - 53.
59 AParty (n 9 above) para 72.
60 AParty (n 9 above) para 74.
61 This is a point lost on the expatriate voting rights lobby which continues to claim that the Richter judgment recognised the constitutional right of all expatriate citizens to vote. See further below.
Chief Electoral Officer (and citizens individually) to ensure the integrity of the registration process and voters’ roll, by constantly reregistering and deregistering voters. The supporting role of citizens in this process is defined in section 9(1) of the Electoral Act, which reads as follows:

A registered voter or person who has applied for registration as a voter and whose name or ordinary place of residence has changed, must apply in the prescribed manner to have that change recorded in the voters’ roll or in that person’s application.

As the Electoral Commission made clear in its heads of argument, the ‘primary statutory duty’ to ensure the integrity of the voters’ roll rests on ‘the voter whose ordinary residence has changed to have that change recorded by the Commission’.62 Registered voters who permanently leave the country must, as a matter of law, inform the Electoral Commission of that fact. Because such voters would no longer comply with the residence requirement in section 8(3) of the Act, the Chief Electoral Officer will have no choice but to remove their names from the voters’ roll.63 In the ordinary legislative scheme of things, an expatriate citizen would lose her eligibility to vote at the same moment that she permanently leaves the country.64 An expatriate South African who failed to comply with her duty as a citizen to deregister as a voter, cannot later turn around and seek to enforce a right to vote on the basis of her registration on the very same voters’ roll whose integrity she had deliberately and unlawfully undermined by becoming a ghost voter. The narrow reading of the second severance order would sanction opportunism in the shadow of the law, which is exactly the opposite of the civic mindedness that the Court seeks to reward and encourage among overseas citizens.65

Secondly, special votes per definition constitute an exception to the general rule and therefore justify a unique application and voting procedure. Under these special circumstances, the mere fact that a registered voter is physically absent from the Republic might provide enough reason to require of that citizen to prove to the satisfaction of the Special Electoral Officer that she is still ordinarily resident in, or only temporarily absent from, South Africa. It cannot be said that it is irrational (arbitrary) or unreasonable (unfair) to expect of all absentee voters to reconfirm their eligibility as voters, given the duties implied by the residence based electoral system as a whole. One could also say that absence from the Republic on election day creates a reasonable suspicion

---

62 Second Respondent’s Heads of Argument para 72.4
63 Section 8(3) read with sec 11(1)(b) of the Act.
64 Whether this should be taken to happen when that person actually leaves the country, or should be deemed to have happened after five years of absence from the country, as in the case of Canada, is beside the point I am trying to make.
65 If only to symbolically underscore this fact, it might be necessary to consider an amendment to the Electoral Act to sanction non-compliance with the statutory duty to reregister or deregister as a voter after every change of ordinary residence.
or presumption that the absentee is no longer ordinarily resident in South Africa. Even if one were to derive a constitutional right to vote from the fact of registration itself, then the temporary absence requirement would constitute a justifiable limitation of that right, given the special circumstances in which the right is exercised and the fact that the requirement was designed as a mechanism to ensure the integrity of the voting process within our residence-based electoral system. Even ordinary voters do not acquire an absolute right to vote upon registration. The right to vote is justifiably limited by a variety of factors. Many of these relate to the logistical side of the process, including the accessibility and opening hours of voting stations. However, some of these factors relate to the integrity of the voting process, like the need to provide positive proof of identity at the voting station. The need to prove temporary absence from the Republic in the case of special votes is of the same category. It is legitimately designed as an internal audit measure to ensure the integrity of the electoral process.

Thirdly, the attempt to reconcile the Richter judgment with the existing residence based electoral system, by distinguishing between the fact of registration and the act of voting, does not stretch very far. It runs into serious difficulties as soon as the question of provincial and municipal elections is broached. The vote of an expatriate citizen on the national level is not tied to any particular voting district. In this instance, the requirement that voters vote per district is largely a logistical and administrative arrangement (as problems with the distribution of ballot papers during the 2009 general elections again underlined). Lack of ordinary residence in South Africa might not itself place a bar on the right of non-resident citizens to vote in this context. The same cannot be said about provincial and municipal elections. As the Minister rightly pointed out, at the provincial level (and even more so at the local level) the voting district in which a voter is registered and ordinarily resident is crucial for the proportional allocation of representatives to a particular provincial legislature or municipal council. Expatriate citizens cannot require a right to vote in provincial and municipal elections simply on the basis of prior registration, without turning the residence requirement of the electoral system into a farce and the fundamental constitutional principle of proportional representation into a fiction. The Constitutional Court itself must have realised this when it found itself unable to resolve the question of special votes at the level of provincial elections, and therefore simply decided to postpone further debate on the issue to another day.

In conclusion, it is simply not possible to escape the complex and contested debate about the nature and benefits of citizenship with a formal appeal to the fact of voter registration alone. The constitutional demerits of

the attempt to distinguish between resident and non-resident citizens as far as voting rights are concerned, must be grounded on other considerations. One such consideration has been suggested by the expatriate voting rights lobby. Myburg called the judgment of O’Regan J a ‘triumph of generosity over mean-spiritedness’ and thereby captured the general spirit of the positive reaction to the case on the internet. This reaction was largely informed by a comment that O’Regan J made about the nature of non-resident or expatriate citizenship as a legitimate form of global citizenship. This remark has been embraced by the expatriate voting rights lobby as the substantive essence, if not the true ratio, of the judgment as a whole. Is this correct? Does the idea of global citizenship, as O’Regan J formulated it, finally provide an explanation for the second severance order and, at the same time, serve as the master principle to govern the construction of citizenship in post-apartheid South Africa?

4.4 The right to vote and global citizenship

The celebrated comment by O’Regan J about global citizenship is contained in a larger passage that reads as follows:

I am influenced by the fact that, as several of the parties noted, we now live in a global economy which provides opportunities to South African citizens and citizens from other countries to study and work in countries other than their own. The experience that they gain will enrich our society when they return, and will no doubt enrich, too, a sense of a shared global citizenship. The evidence before us, too, shows that many South African citizens abroad make remittances to family members in South Africa while they are abroad, or save money to buy a house. To the extent that citizens engaged in such pursuits want to take the trouble to participate in elections while abroad, it is an expression both of their continued commitment to our country and their civic-mindedness from which our democracy will benefit.

The first thing to note is that this passage was written with resident citizens in mind, who will soon be returning to the Republic after a temporary sojourn abroad. However, Myburgh suggested that this passage was equally applicable to non-resident citizens who had permanently left the country with no fixed intention to return. According to him, the importance of the passage lay in the ‘message it conveys to South Africans living abroad’. According to Myburgh, O’Regan J was telling emigrants that their continued commitment to the country and civic mindedness will not go unrewarded and thus revealed a desire on the part of the Court to ‘foster a continued sense of belonging, and help keep alive the desire to return’ among expatriate citizens. Macdonald similarly read the passage as

68 Richter (n 9 above) para 69. The passage is quoted with approval by Myburgh.
an attempt to reach out to an expatriate community which has been forced into exile by crime and the affirmative action policies of the ANC government, but which remained patriotically South African at heart. According to Macdonald, granting expatriates the right to vote will ‘make our scatterlings feel like they are still part of the country and part of its future. Denying them the right risks alienating them and causing them to lose interest in their native land and be lost forever.’

Whether these commentators get the point that O’Regan J was trying to make or not, they completely miss an important point about democratic citizenship. The right to remain engaged in the political processes of a state and to have a say in the making of its laws cannot be acquired as a reward for continued national patriotism or entrepreneurial success, it can, democratically speaking, only originate as a precondition of being subject to the jurisdiction of that state, its courts and its laws. A fundamental virtue of democratic citizenship is precisely that it places limits on the way in which economic power, mobility or agency translate into political power. As Rubio-Marín argues, we should therefore be wary not to express the economic activities of expatriates as a form of economic citizenship and then use that concept as the basis for the allocation of political rights. The link that O’Regan J seems to establish between economic productivity on the international stage and good citizenship, threatens to completely reduce political action to neo-liberal market interaction. It thus all but totally instrumentalises the right to vote, at the cost of a wide range of more deliberative or republican understandings of political participation and action. Voting rights are approached as simply another means to promote a citizen’s individual or collective interests. Liberals have always

71 A Ong ‘Mutations in citizenship’ (2006) 23 Theory, Culture and Society 499 has recently warned that the celebration of neoliberal market values, like flexibility, mobility, and entrepreneurialism as ideal qualities of citizens, seriously threaten the republican ideals of political equality and democratic freedom.
72 See further F Michelman ‘Conceptions of democracy in American constitutional argument: Voting rights’ (1989) 41 Florida LR 443. Michelman distinguishes between two conceptions of democratic politics that are implicit in US American voting rights cases. He calls these deliberative politics and strategic politics (447). He also distinguishes between two reasons why people might value enfranchisement. He calls these instrumental reasons and constitutive reasons. In terms of the former, people use the right to vote as an instrument to further their private interests. In terms of the latter, people make use of the right to vote to constitute and constantly transform their individual and communal identities in a public dialogue with other citizens (451). Ostensibly to ensure the dialogical or deliberative nature of politics, republicans have traditionally supported the introduction of a range of voter qualifications (including literacy tests). Michelman takes up the challenge to rid republicanism of this tainted voting rights tradition, without giving up on the value of dialogical, as opposed to instrumental, politics. In this context, Michelman accepts bona fide residence as a constitutionally-permissible franchise prerequisite in US American law (469 - 480). Bona fide residence can withstand close constitutional scrutiny because it is necessary to preserve the (dialogical) ideal of political community or the value of democracy
thought that politics and political interests could be taken care of through representation (just as economic interests can be a taken care of through representation). The interest-based model of democracy already implies absentee membership of, or participation in, politics. It is therefore not surprising that this undemanding model of democracy can easily be extended to include absentee voting rights in favour of non-resident citizens who have or take an interest in South Africa.

A comparison between the South African and Mexican attempts to embrace migration as a strategy of development might be instructive in this regard. Kim Barry claims that the Mexican case indeed provides a paradigmatic example of the fact that the extra-territorial relationship between nation states and their emigrant communities are being reconstructed under the pressure of globalisation. States are increasingly willing to tolerate and even encourage migrant identities with multiple allegiances and loyalties. Gone is the romantic association of nationality with an exclusive patriotism to the state and nation. In line with the global trend towards the recognition of dual citizenship, Mexicans who have naturalised abroad are now allowed to retain or to restore their Mexican citizenship or nationality. Mexico’s new self-conception as an emigration state thus includes a celebration of the fact that the Mexican nation extends beyond the territory of the Mexican state and its jurisdiction. On the basis of this diasporic self-conception, greater economic and cultural ties are actively pursued between emigrant communities and Mexico as homeland. However, the point that Barry seeks to underscore is the following: Although the economic and cultural engagement of emigrant communities are increasingly welcomed, their ‘direct participation in the national political community generally is not’. Even in a country where the popular and official view of expatriates has shifted from traitors to heroes, that reorientation has not translated into the recognition of expatriate voting rights. The opposite might even have been the case. The reorientation has resulted in a far clearer distinction between the legal ties that a person might have with a state, the economic and cultural ties she might have with a nation, and the political ties she might have with a political community. To accommodate these distinctions, Barry suggests that we restrict the term citizenship to this last-mentioned category and distinguish more demandingly than is generally the case between nationality (or formal legal citizenship) and citizenship (or active participatory citizenship).
Barry’s discussion of the Mexican case suggests that we need a more robust theory of democratic citizenship as political membership, stripped from its cultural, economic and international law dimensions. The challenge to develop such a theory has been taken up by Lôpes-Guerra. Lôpes-Guerra points out that the argument for voting rights, based on the ongoing economic contribution of the expatriate community to their country of nationality, is but one of several desert based arguments for the right to vote. She rejects the idea that citizens can acquire voting rights on any desert based consideration. Democracy requires that everyone who is directly subject to the territorial jurisdiction of the law should have a say in the making of that law. The crucial question is thus not whether expatriate citizens remain patriotic or economically productive Mexicans or South Africans, as O'Regan J has been read to suggest, but how expatriate citizens who are no longer directly subject to the jurisdiction of the law can rightfully claim a democratic right to have a say in the making of the law?

The difference between the diasporic nationalism or national patriotism defended by Myburg and Macdonald and the constitutional patriotism defended by Rubio-Marín and Lôpes-Guerra is sharply drawn. The idea of constitutional patriotism is generally associated with the work of Habermas. Habermas argued during the early 1990s that the task of the next generation of critical theorists was to employ a discourse theory of societal rationalisation in order to free democratic republicanism (or citizenship) from the domination of administrative state power and the mythology of the nation. In contrast to the instrumentalist and ethical understandings of citizenship, Habermas favoured a discourse theoretical understanding in which the political community of democratic citizens was constituted by a network of egalitarian relations of mutual recognition that were called into being by the discursive medium of law itself. In this regard, he felt encouraged at the time by the remark of the German Constitutional Court, in an important voting rights case of the late 1980s, to the effect that

the idea of democracy, and especially the idea of liberty contained in it, implies that a congruence should be established between the possessors of democratic political rights and those who are permanently subject to a specific government. This is the proper starting point.

The liberal expatriate lobby presents us with a theory of the law, democratic politics and voting rights in which this insight is almost completely forgotten. The basic idea underlying expatriate voting rights is that certain citizens can insulate and remove themselves from the jurisdiction of the law, without affecting their relationship with the rest of

77 C Lôpes-Guerra ‘Should expatriates vote’ (2005) 13 Journal of Political Philosophy 216
227.
78 J Habermas Between facts and norms: Contributions to a discourse theory of law and democracy (1996) 499.
79 Habermas (n 78 above) 509.
the political community who are living under the law. The right of expatriates to vote therefore means the right to impose obligations on resident South Africans without having to assume those very same obligations. Whether or not this right is implied by collective membership in a nation, it is not implied by collective membership in a democratic community in which democratic accountability among citizens is the defining feature.

There might be good reasons in international law why South Africans may not be stripped of their formal citizenship (nationality) once they have permanently taken up residence in another country. However, there is no good reason of democratic theory why such non-resident citizens may not be stripped of their right to vote. Many democracies (like Canada) allow non-resident citizens to retain the right to vote for a predefined period of time only, after which expatriate citizens are automatically stripped of the right to vote. Again it is useful, as Barry has suggested, to more carefully distinguish in this regard between nationality and citizenship, as the two concepts relate respectively to membership in a state for the purposes of international law and membership in a bounded democratic political community for the purposes of election law. Membership in the latter (citizenship in the narrow sense of the term) should be determined with reference to the value of democratic accountability: those who are directly subject to the jurisdiction and violence of law should have a say in the making and administration of the law; those who are not, like expatriate citizens who permanently live and work outside the jurisdiction of the state, should not.

The relationship between voting rights and the value of democratic accountability implied in the territorial jurisdiction of the law, suggests a close relationship between voting rights and residence. Benhabib has recently identified the rise of new forms of post-national street-based citizenship (or denizenship) as a significant reaction to the twilight of nation-state sovereignty. Post-national citizenship refers to the rise of political activism on the part of non- and post-nationals living together in multi-cultural and ethnic inner-city neighbourhoods, who come together around issues like women’s rights, secondary language education, environmental concerns, representation on school boards and migrant employment. The new forms of ‘urban activism’ described by Benhabib can be understood as a perfect example of the ideal of street democracy that I defended on a number of previous occasions. Benhabib’s point is

---

80 Benhabib (n 4 above) 263.
81 This new urban activism and politics need not be centred on voting rights and integration into representative law-making institutions. Benhabib favours the institutional recognition of post-national denizenship through voting rights, as has already taken place at the local government level in the European Union. There are obvious limits to any focus on voting rights as an entry into transformative post-nationalist and post-national politics. The focus on voting rights tends to reduce political action to voting, and thus immediately folds the potentially radical idea of a
that post-national or urban citizenship transcends the old distinction between national citizens and non-citizens, and relocates democratic citizenship in local residence and active everyday life on the streets of the world’s great migration cities.83

Today we are caught not only in the reconfiguration of sovereignty but also in the reconstitution of citizenship. We are moving away from citizenship understood as national membership increasingly towards a citizenship of residence which strengthens the multiple ties to locality, to the region and to transnational institutions. [...] This new modality decouples citizenship from national belonging and being rooted in a particular cultural community.

The value of democratic accountability implied in the territorial jurisdiction of the law requires, as a starting point, that democratic citizenship be tied to ordinary residence. It requires, as Agamben argues, that citizens of nations increasingly turn themselves into denizens of cities.84

5 Conclusion

A full defence of a post-apartheid theory of denizenship along these lines will have to wait for another occasion. On this occasion I simply considered whether the recent voting rights cases of the Constitutional Court create any insurmountable difficulties for any such theoretical defence. My conclusion is that they do not. While the Richter case rejects any distinction between resident and non-resident citizenship for the purpose of absentee voting rights, the judgment ultimately fails to provide any constitutional justification for this rejection. On closer analysis, the Court’s tentative appeals to voter registration and global citizenship cannot serve as ground for that rejection. Given the Constitutional Court’s own reluctance in the AParty case to engage with the complex constitutional connections between residence and political participation, the second severance order of Ebersohn J should never have been confirmed in the

democratic civil society back into the representative law-making and law-applying institutions of the state. Politics is then reduced to law making and litigation. This point is forcefully argued by B Honig ‘Another cosmopolitanism? Law and politics in the new Europe’ in S Benhabib Another cosmopolitanism (2006) 102 against Benhabib’s law-centred approach to politics or ‘democratic iteration’. This is also the basis of Karin van Marle’s critique of Christof Heyns’s struggle theory of rights (see K van Marle ‘Lives of action, thinking and revolt: A feminist call for politics and becoming in post-apartheid South Africa’ in W le Roux & K van Marle Post-apartheid fragments: Law, politics and critique (2007) 34, 41).


83 Benhabib (n 4 above) 262.

84 Agamben (n 3 above) 22 - 24.
Richter case. As the matter stands, the Electoral Act accepts ordinary residence as the key to the political participation of citizens. While the Richter case might have unsettled the accepted link between residence and democratic citizenship, it has certainly not broken this link.

This is unfortunately not how the expatriate voting rights lobby has chosen to read the Richter judgment. The Centre for Constitutional Rights, in particular, has entered into talks with the Minister of Home Affairs about the introduction of legislative amendments that would bring the Electoral Act ‘in line with’ the Richter judgment by allowing all non-resident South African citizens to vote in future national and provincial elections.85 This campaign unfortunately continues to be based, not only on a misreading of what the Richter case ruled, but also on the constitutional merits of that ruling. Once the distinction between resident and non-resident citizens that is drawn in section 33(1)(e) and the rest of the Electoral Act is understood as constitutionally mandated by the value of democratic accountability, it becomes possible to radicalise the idea of resident nation state citizenship into a form of post-national and nationalist denizenship. In the light of what was said above, if any legislative reform to the Electoral Act is needed after the Richter judgment, it is not to allow non-resident citizens to register and vote in the next election, but to enable non-citizen residents to register and vote in the next municipal elections of the cities and streets where they ordinarily reside.

85 De Havilland (n 55 above) continues to (misleadingly) claim that the Richter judgment recognised the constitutional right of all citizens to vote, regardless of where they ordinarily reside: ‘The Constitutional Court’s ruling in the matter of Richter v Minister for Home Affairs was only a limited victory for those who sought the vote for South Africans overseas. In effect, the right to vote was assured only for South Africans who were already registered to vote - and then only in the national and not the provincial elections. This did not mean that the Court did not accept that all South Africans living overseas have a right to vote in all elections, but that there was simply not sufficient time to make the necessary arrangements before the elections’ (my emphasis). The Centre for Constitutional Rights has driven their campaign for legislative reform on the mistaken claim that there is a constitutional obligation on the executive and Parliament to introduce and to adopt the necessary legislative amendments to give effect to the Richter judgment: ‘The clear requirement is that the Parliament and the executive should adopt the legislation and regulations that will enable all South Africans overseas to vote in future elections, well before the next election takes place.’ Unfortunately, it would appear as if the Minister of Home Affairs is now also approaching the matter on this basis. As De Havilland reports: ‘Last month the Ministry of Home Affairs informed the Centre Foundation that the Department is considering legislative amendments to bring the relevant legislation “in line with the Richter case and other related cases heard before the ConCourt just before the elections earlier this year”’. These amendments would be dealt with in conjunction with the Independent Electoral Commission ‘as it is the entity that implements the Electoral Act’.
Wessel le Roux's paper argues for new conceptions of citizenship. His starting point is a change in the old Westphalian world order that must give way to new forms of empire, transnationalism, postnationalism and cosmopolitanism. He critically reflects on the Constitutional Court decision in Richter. Willem Richer argued that he should be allowed to vote in the general elections of 2009 while working abroad. The Electoral Act allowed a special vote for voters absent at the time of elections for the purposes of a holiday, a business trip, attendance of a tertiary institution or an educational visit, or participation in an international sports event, but not for someone working abroad in a private capacity – in other words, the person is not working abroad for diplomatic or government purposes.

As Le Roux explains, this case disclosed the issue of who should or should not be allowed to participate in the political life of the country through the voting process – resident citizens, non-resident citizens and/or non-citizen residents. Without going into all the technical details already explained by Le Roux, the implication of the Constitutional Court’s decision is that not only voters who are temporarily abroad for private work should be allowed to vote, but also those who are permanently abroad. In doing this, the court negated the residence-based approach that stands at the heart of the current Electoral Act. It is particularly this aspect of the judgment that troubles Le Roux.

I will focus on only a few of the issues raised – Le Roux’s main concern is the expansion of voting rights to non-citizen residents as a way of allowing migrant populations to be politically active in the place where they are not merely working but, in fact, living. Related to this is a
particular understanding of and support for politics, democracy and citizenship, one that breaks with the liberal view that supports an instrumentalist conception of politics and democracy. Le Roux points out that the only concern in the liberal view is the relation between the individual and the state; voting rights are important merely for the protection of an individual’s interests. According to this view political and economic interests could be protected sufficiently by representation. Liberal politics would be content with the idea of non-resident voting. Le Roux echoes the critique of many others that this is an instrumental concept of politics, democracy and citizenship and one that the South African Constitution and Constitutional Court rejected.

The model embraced by the Constitutional Court and supported by Le Roux is one of participatory democracy. In this understanding democracy and politics means much more than going to the polls once every four years. This version of democracy is linked to a republican notion of politics, where collective efforts, public deliberation and some notion of a common good stand central. But Le Roux highlights another aspect, namely the idea of democratic accountability. Politics is not merely about maximising individual interests, but also about being accountable, having certain duties and obligations. This argument brings him to make the following statement: ‘Only citizens who are directly subject to the jurisdiction of the South African courts should be able to vote.’ Non-resident citizens could thus be ‘stripped of their right to vote’. To be clear:

[C]itizenship in the narrow sense of the term should be determined with reference to the value of democratic accountability: Those who are directly subject to the jurisdiction and violence of law should have a say in the making and administration of the law; those who are not … should not.

The element of accountability relates to the necessity of residence, and further the issue of being subject to the law. Expanding voting rights to non-citizen residents in this view must mean a simultaneous limiting of voting rights of non-resident citizens.

Up to this point I have highlighted the following from Le Roux’s paper: a call for the expansion of voting rights that implicates a rejection of a liberal idea of politics and democracy and a support for a participatory idea. Relating participatory politics to obligation, paradoxically leads to a limiting of voting rights, in other words in the same gesture that we expand voting rights to non-citizen residents, we limit the voting rights of non-residents or rather strip non-resident citizens of their voting rights. The distinction between formal citizenship (nationality) and active participatory citizenship is used to explain this move. Le Roux also recalls the distinction made by Habermas and followed by, amongst others, Benhabib between national patriotism and constitutional patriotism. I elaborate on the notion of constitutional patriotism below.
Referring to arguments justifying voting rights for non-residents on the basis of their economic contribution, Le Roux argues that the main concern is not whether expatriate citizens remain patriotic or economically productive, but that expatriate citizens who are no longer subject to the jurisdiction of the law should not have a democratic say in the making of the law. The law and the political life enabled or disabled by law is thus crucial for the view that refuses non-resident voting rights.

It is this embrace of law, and in particular of the Constitution, that I will focus on now. My concern is exactly the centrality given to the law and to the Constitution; politics, democracy and citizenship are directly related to and enabled by law. Although I support the notion of expanding voting rights to non-citizen residents and agree with the critiques levelled against a liberal, instrumental notion of politics and democracy, I am unsure about the emphasis on residence as the yardstick for participation and, more than that, the concession to law and constitutionalism as a vessel for community, politics and democracy.

Christodoulidis cautions against what he calls the ‘optimism’ inherent in an endorsement of a republican understanding of politics, democracy and citizenship. His claim is specifically against the republican belief that the law and in particular constitutionalism can serve as a basis for politics and the community. Although the vision of an active citizenship is promised because of the containment inherent in law and constitutionalism, the ‘politics of civil society’ will be inhibited and the ‘acquisition of social and political identity’ obscured. Christodoulidis explains constitutional patriotism as follows: ‘What Habermas has termed “constitutional patriotism” pronounces its ties to both active participation and membership in the state (and consequently the nation due to the historically contingent connection between state and nation).’ What is significant about Christodoulidis’ argument is that he exposes how both models are connected to law. The liberal model and the republican model have a shared premise, namely law. ‘The existence of the constitutional framework underpins both competing (liberal and republican) positions because neither transcends the language of citizenship and rights and both seek their anchorage, ultimately, in law.’ The differences of course should be noted, what is important for the liberal is ‘accommodation of interest’, ‘a one-way process of feeding his/her contribution’ and what is important for the republican is ‘inter-subjectively shared praxis’, ‘a feedback’. In the republican view, law provides the substantiation for popular sovereignty.

---

2 Christodoulidis (n 1 above) 178.
3 Christodoulidis (n 1 above) 179.
4 Christodoulidis (n 1 above) 180.
5 As above.
through the constitution, by providing a constitutional/political home for participation and deliberation.\(^6\)

Bankowski and Christodoulidis, within the context of debates on the issue of citizenship in the European Union, argue for citizenship based on ‘participation’ and not any form of ‘received membership’.\(^7\) For them, ‘a community’s constituency should not be taken as a given as when membership is its precondition’. A community should in fact contest its political space, and, more than that, is generated by treating its political space as contestable, thereby ‘postponing its fixity’.\(^8\) They attempt to create a position between ‘bounded’ (linked to nation) and ‘unbounded’ (linked to globalism) notions of citizenship by putting forward the notion of an ‘essentially contested’ constituency. According to the authors in the contestation over constituency ‘community comes about’.\(^9\) In this context, they invoke the metaphor of home but strongly argue for a conception of home that is fluid and constantly changing.

I raised a concern above, that Le Roux, by supporting the notion of constitutional patriotism, is making law and the Constitution too central to citizenship, political participation and democracy. I agree with his suggestion of expanding voting rights to non-citizen residents and agree with his critique of a liberal understanding of politics and democracy, resulting in a support of negative freedom and the notion of fixed identity. However, my sense is that the trust in constitutionalism as a vessel for politics reduces the kind of politics, political community and identity that Le Roux himself supports and by institutionalising democracy (the street) limits exactly the expansion of voting rights to non-citizen residents. We can only recall the case of Khosa and the Constitutional Court’s inclusion of permanent residents (non-citizens) in the social security allowance of the state, but not of other people finding themselves in need of social security, albeit not legally declared as permanent residents. We could ask why voting rights of permanent residents are, in Le Roux’s conception, limited to local elections, why can it not be expanded to national elections. One line of argument could be that the reasoning for the expansion is residence, that political rights are linked to a material political community that permanent residents must be allowed to have a say in local issues.

Following the Arendtian line that Le Roux follows to a certain extent, we must recall her distinction between labour, work and action and that, for her, action is a space where people appear to each other, not where social issues are elevated to politics. Of course, Arendt has been criticised for this view: How can social and economic issues not be seen as political

\(^6\) As above.
\(^8\) As above.
\(^9\) Bankowski & Christidoulidis (n 7 above) 89.
issues in present times? But then Le Roux is making a similar distinction between economic activities, issues of identity and politics in his argument as to why South African citizens living abroad, contributing to the South African economy, and identifying themselves as South Africans, should not be allowed to vote. Politics and democracy for Le Roux are separate from issues of economics, culture and identity. My question is whether we can make such strict distinctions, are not these issues more fluid – particularly if the claim is for identities that are continuously constructed. My sense is that Le Roux’s argument for expanding voting rights to non-citizen residents instinctively ‘gets’ the impossibility of such a distinction.

Le Roux underscores the importance of residence, and makes thereby a certain understanding of space and time central to his argument. As with the issue of identity we might consider a less rigid understanding of time and space that I think would be more in line with Le Roux’s conception of the street. My main contention is thus that Le Roux by relying so much on law and the Constitution undercuts his own vision of an alternative citizenship. Historically, property rights and ownership were seen as central to citizenship and voting rights. Le Roux is clearly trying to question this. However, by relying so much on law and residence, he is almost institutionalising living, life itself. It is not ownership or a vested right in property that is connected to political rights, but the normative claim of living (and accepting?) the norm of the law.

I referred above to Bakonwski and Christodoulidis’s reliance on the metaphor of home and their support of a notion of home that is fluid and constantly changing. Du Toit, within the context of rethinking feminine subjectivity, relies on Young’s argument for an alternative vision of home.10 Women have been linked to a certain understanding of home – an understanding that restricts women to the traditional role of caretaker and nurturer of husbands and children. Du Toit raises a crucial question: ‘Where do women’s homes reside, in distinction from the homes that women create for men?’11 The argument is that women as homemakers find themselves homeless and rootless and that these conditions keep women in a place of subordination. Young puts forward a vision of home that could afford women a sense of belonging and that breaks with the traditional conception. The notion of a constitution is quite often referred to as providing a political home to the citizens and in most aspects residents of a country. However, many South Africans (residing and non-residing) do not identify with the Constitution and find themselves politically homeless. If the concern is with a notion of politics and democracy that could provide a notion of home to South African citizens and non-citizen residents my concern again is that constitutional

11 Du Toit (n 10 above) 26.
patriotism instead of expanding political rights and invigorating democracy will not succeed.

Du Toit relies on a phrase from a story by Jeanette Winterson: The goddess Artemis, uncomfortable with the traditional trappings of womanhood, persuades her father to allow her to be a hunter. Du Toit observes that Artemis ‘in the process … quickly discovers that the real challenge of freedom has more to do with spiritual strength (and learning to live with all one’s various selves) than with physical ability or movement’.\textsuperscript{12} She makes herself ‘at home’, lives in a shack with her dogs, a temporary home, rather than living as someone’s wife, mother or daughter creating home for them. The god Orion, upset because of her rejection of him, destroys her home, kills her dogs and rapes Artemis.

Artemis kills Orion with a scorpion and discovers another notion of home; she discovers that it is possible to ‘hang in space supported by nothing at all’. Du Toit interprets this phrase as a rejection of nostalgia for the perfect, timeless and maternal home. For Du Toit, this story opens the question of whether it is possible to think of a feminine identity that neither affirms masculine identity nor affirms notions of nothingness. The phrase hanging in space supported by nothing at all captures for her the need to have a space, but the lack of a fixed foundation. My question is: To what extent can we think of notions of citizenship, political subjectivity and democracy that do not amount to an eclectic postmodern anything goes – as some notions of cosmopolitanism might entail – but that is also not limited to life under the law and the Constitution?

\textsuperscript{12} Du Toit (n 10 above) 30.
At the heart of the South African Constitutional order lies an enterprise that has been characterised as 'transformative constitutionalism'. This entails that the Constitution in South Africa was not designed simply to entrench the status quo; rather, it was enacted for the purpose of fundamentally transforming society.

However, the notion of transformation or change alone tells us very little: in order to comprehend what the Constitution is designed to achieve, it is necessary to have an understanding of the features of the past that must be discarded and to have a vision of the type of future towards which South African society should be heading. This point is enshrined in the Preamble of the Constitution which outlines its aim as being 'to heal the divisions of the past and to establish a society based on democratic values, social justice and fundamental human rights'.

The Constitution also places three central values at the core of the society it is designed to create: human dignity, the achievement of equality and the advancement of human rights and freedoms.

--


2 In Klare’s words, the enterprise is one of ‘inducing large-scale social change through non-violent political processes grounded in law’ Klare (n 1 above) 150.

3 Preamble to the Constitution of the Republic of South Africa 1996.

4 Constitution sec 1.
These statements give us a sense of the mission and purpose of the Constitution, a purpose that aims at the highest ideals for South African society. The first part of this article considers the historical backdrop to the Constitution and some of its key provisions to argue for at least one fundamental principle or *grundnorm*\(^5\) that can be said to underlie the manner in which constitutional interpretation is to be effected. The principle in question requires that there be no arbitrary discrimination against any particular group without a justifiable reason. This principle, it shall be argued, applies not only to questions of race but to all forms of arbitrary discrimination including that which is based on such grounds as gender, age and sexual orientation. When making sense of the application of this principle, the question arises as to whether arbitrary discrimination can occur not only between members of the human species but between different species. It is argued that there are no good grounds to focus arbitrarily on human beings alone and so to exclude non-human species from the application of this *grundnorm*. If this is so, the question arises as to how the Bill of Rights can be interpreted so as to extend its protections to non-human animals in seeking to realise an ethos of non-arbitrariness.

This raises the problem that there are no express provisions granting protection to animals despite the fact that their fundamental interests can be severely impacted upon by human beings. If non-human animals are the proper subjects of our moral concern, and the Constitution (and wider legal order) provides them with no protection, then it appears that there is a grave omission at the heart of our constitutional order. A constitutional amendment would be the most desirable route to remedy this omission.\(^6\) However, in the absence of such an amendment, I attempt to develop in the second part of this article an interpretation that, in light of the fundamental *grundnorm* underlying the Constitution, abjures discrimination on grounds of species and extends the protections of the Bill of Rights to non-human animals. First, I consider the provision governing the application of the Bill of Rights and why the designation ‘everyone’ in certain rights can be interpreted to include non-human animals. Secondly, I consider the value of human dignity and whether this indicates that the Bill of Rights was designed only to provide protections for human beings. I argue that the historical context of the Constitution provides strong reasons why the protection of human dignity was included as a core value: The appellation ‘human’ was not designed, however, to exclude the recognition of other forms of dignity or worth. The fundamental *grundnorm* of non-discrimination would entail that the Constitution must be taken to recognise that all sentient creatures have worth rather than focusing only on those who belong to a particular

---

\(^5\) I use the notion of a *grundnorm* in the sense of being a fundamental principle underlying and underpinning the very legitimacy of the constitutional order in South Africa.

\(^6\) The current political climate would not seem favourable to such an amendment. The interests of animals rarely make it on to the agenda of parliament and, even in relation to environmental legislation in South Africa, tend not to take centre stage.
Does transformative constitutionalism require the recognition of animal rights?

grouping – namely, *homo sapiens*. The dangers of restricting the recognition of dignity to human beings alone will be illustrated by reference to an important case that arose in Germany. Thirdly, the value of equality is considered. An argument is provided to the effect that the Constitution should not simply recognise that members of other species have some worth in their lives: rather, the presumption should be in favour of recognising that they have *equal* worth. Such a presumption does not, however, entail identical treatment and the protections accorded non-human animals should indeed vary with their capabilities.

If the arguments of section two are convincing and animal rights are recognised, the circumstances under which they can be limited become of key concern. Indeed, this raises a major problem in that South African society, as it stands, does not seem ready to acknowledge fully the implications of the recognition of animal rights which would, for instance, require a general ban on eating meat. It is unlikely that many individuals in South Africa would support such a position and, politically, it seems unlikely that the full implications of animal rights will be recognised in law. It is thus necessary to consider ways in which to limit such rights in the shorter-term without rendering them meaningless. Part III seeks to consider two facets of the Bill of Rights that can assist in this process: the notion of ‘progressive realisation’ and the general limitations enquiry. An important case in which judges grappled with these complex issues in Israel will be discussed. The article concludes by suggesting that the logic underlying human rights in fact provides the foundation for recognising animal rights. South African society will only truly be transformed once it realises the full implications of the *grundnorm* at its heart and extends the protections of its Constitution to all who deserve them irrespective of the species to which they belong.

1 The historical imperative: No arbitrary discrimination

1.1 History and constitutional interpretation

The project of transformative constitutionalism has been characterised by Klare as involving a ‘long-term project of constitutional enactment, interpretation, and enforcement committed … to transforming a country’s political and social institutions and power relationships in a democratic, participatory and egalitarian direction’. What Klare’s statement foregrounds is the importance of understanding the changes required by the Constitution and the values which should guide this transformation. Any characterisation of the changes required by the Constitution, however, must require reference to an understanding of the injustices that it represents a reaction against. These points have been underscored by the Constitutional Court in numerous judgments, one of the most celebrated of

---

Klare (n 1 above) 150.
which is *S v Makwanyane*.\(^8\) The majority judgment of Chaskalson P recognises the importance of adopting a ‘purposive and generous’ approach to interpretation that ‘gives expression to the underlying values of the Constitution’.\(^9\) Such an interpretation must construe rights in their context, ‘which includes the history and background to the adoption of the Constitution, other provisions of the Constitution itself and, in particular, the provisions of chapter three of which it is part’.\(^10\) These statements were confirmed by a number of judges: Mokgoro J, for instance, stated that constitutional interpretation requires reference to ‘a system of values extraneous to the constitutional text itself, where these principles constitute the historical context in which the text was adopted and which help to explain the meaning of the text’.\(^12\)

The interpretive method of the court thus suggests the importance of history in construing the value-based choices that judges are to make when interpreting the Constitution. What is necessary, at a minimum, is to provide some substantive understanding of the nature of the injustices of the past. In so doing, we come to understand what moral and political failures must be avoided in the future.

Understanding the nature of the injustices of the past is a complex task and can be prone to over-simplification.\(^13\) I thus do not seek in this article to provide an exhaustive analysis thereof. However, I wish to focus on a particularly egregious element of South African history – something I would suggest is relatively uncontroversial – and then seek to draw out a central moral principle that must guide constitutional interpretation in the future. The application of this principle to non-human animals will then be considered.

### 1.2 Apartheid and the grundnorm of non-arbitrariness

One of the key features of South African history has been a denial of equal treatment to black people who were treated in an inferior manner simply because of their race.\(^14\) Indeed, in providing some historical context to the adoption of the Constitution, the Constitutional Court has stated that
‘[f]rom the outset the country maintained a colonial heritage of racial discrimination: In most of the country the franchise was reserved for white males and a rigid system of economic and social segregation was enforced.’15 In describing the central elements of the apartheid system, the Court states that ‘[r]ace was the basic, all-pervading and inescapable criterion for participation by a person in all aspects of political, economic and social life’.16

Mahomed J (and many of the other judges) recognised this as one of the key facets of South African history that the Constitution was designed to take us away from

the contrast between the past which it repudiates and the future to which it seeks to commit the nation is stark and dramatic. The past institutionalised and legitimated racism. The Constitution expresses in its preamble the need for a ‘new order … in which there is equality between … people of all races’ … The past was redolent with statutes which assaulted the human dignity of persons on the grounds of race and colour alone; section 10 constitutionally protects that dignity. The past accepted, permitted, perpetuated and institutionalised pervasive and manifestly unfair discrimination against women and persons of colour; the Preamble, section 8 and the postamble seek to articulate an ethos which not only rejects its rationale but unmistakenly recognises the clear justification for the reversal of the accumulated legacy of such discrimination.17

It is important to analyse in a little more depth, the reasons why we regard discrimination on grounds of race as so repugnant. Perhaps the key moral wrong is that it separates out the treatment afforded to a person from any morally relevant feature of their being: for instance, black people were subject to inferior education not on the basis of any lack of ability but simply because they were black. The system of apartheid thus enshrined into law a form of moral arbitrariness that separated out the treatment of an individual from their individual capacities.

As Rachels points out, ‘[f]acts about people often figure into reasons why they may not be treated in this or that way. Adam may be ejected from the choir because he can’t sing. Betty may be given Prozac because she is depressed. Charles may be congratulated because he has just gotten engaged … Notice, however, that a fact that justifies one type of treatment may not justify a different type of treatment: unless something unusual is availability of economic opportunities.’ I do not seek to deny that other factors contributed to apartheid such as the economic motivation for cheap labour: Rather, what I wish to assert here is that a central feature of apartheid was institutionalised racial discrimination which, at least to some extent, was based upon an ideology of white supremacy. See De Vos (n 13 above) 14 for a critique of construing the evils of South African history too narrowly.

16 Certification judgment (n 15 above) para 5.
17 Makwanyane (n 8 above) para 262.
going on, we could not justify giving Betty Prozac on the grounds that she can't sing or throwing Adam out of the choir because he has become engaged.\footnote{18}

One of the most fundamental moral deficiencies of apartheid was thus to treat individuals in a prejudicial and inferior manner on the basis of characteristics that provided no justification for such differential treatment. If the new constitutional order is to represent a fundamental shift in this regard, then it must set its face against this serious moral failing. Such a shift, however, cannot only focus on discrimination on grounds of race: To do so would reproduce the very arbitrariness that is meant to be discarded. Instead, the focus must be on rejecting all forms of treatment which arbitrarily discriminate against individuals. Thus, understanding the historical context in which the Constitution occurs, suggests that there is a fundamental moral principle or \textit{grundnorm} that must guide the interpretation of the Constitution in any transformed legal order in South Africa: 'Do not subject any individual to differential (and, particularly, prejudicial) treatment without any adequate justification.'\footnote{19}

When we consider the text of the Constitution itself, there are two key features which appear to give expression to this \textit{grundnorm}. Section 1, outlining the founding values of the Republic, refers to the supremacy of the Constitution and the rule of law.\footnote{19} The rule of law involves a number of elements: One of its key facets though is a commitment to general principles that apply across the board. Continued \textit{ad hoc}, arbitrary treatment of individuals would not be acting according to law but would be its antithesis.\footnote{20}

A famous example of such a violation of the rule of law occurred in South Africa: A law was passed by parliament to allow for the renewal of the internment of Robert Sobukwe simply on the basis of a discretion exercised by the Minister of Justice.\footnote{21} The clause was designed around Sobukwe and he was the only person to have his internment extended under it. The Constitution of South Africa deliberately sets itself against such laws requiring that any justifiable limitation of a right can only occur through a law of general application.\footnote{22} This requirement is rooted in the rule of law.\footnote{23} The Constitutional Court has on several occasions confirmed the fact that the rule

\begin{footnotes}
\item Sec 1(c) of the Constitution.
\item See L Fuller \textit{The morality of law} (1969) 46 - 49 for the argument that for law properly to exist, it must meet certain criteria of generality. See also J Finnis \textit{Natural law and natural rights} 270 - 271.
\item General Law Amendment Act 37 of 1963.
\item Sec 36(1) of the Constitution.
\end{footnotes}
of law requires that public power must not be exercised arbitrarily or manifest ‘naked preferences’.24

The second element of the Constitution that most clearly expresses the grundnorm in question is the founding value of ‘achieving equality’ (section 1(a) of the Constitution) and, perhaps, most importantly the right to equality in section 9 of the Bill of Rights. It is no accident that equality is the first right enumerated there and the context of South African history provides a clear understanding as to why this was so. The equality right begins in section 9(1) with a requirement that ‘everyone is equal before the law and has the right to the protection and benefit of the law’. Section 9(2) allows for differential treatment such as affirmative action that is designed to ensure that previously disadvantaged groups are able to achieve equality given their unequal starting points. Section 9(3) prohibits unfair discrimination by the state on a number of grounds that include race, sex, sexual orientation, age, religion, and several others. Section 9(4) prohibits unfair discrimination by individuals on those same grounds.25 Importantly, the grounds upon which unfair discrimination is prohibited are not exhaustive and other grounds can be added.26 The Constitutional Court has, for instance, recognised that non-citizens may also not be subject to unfair discrimination.27 This is important in that it involves a rejection of the idea that unfair discrimination may only occur on particular specified grounds: rather, there is an affirmation and confirmation of the grundnorm I have identified, which is that the Constitution sets itself against all forms of arbitrary discrimination against individuals.

1.3 Arbitrariness and non-human animals

If all forms of arbitrary discrimination are proscribed by the Constitution, the question then arises as to whether this applies only to human beings or between human beings and other species as well. In particular, the question arises whether failing to recognise that non-human animals have fundamental rights constitutes a form of unfair discrimination against them. In other words, can the grundnorm I have identified and that

24 See Pharmaceutical Manufacturers Association of SA: In Re Ex Parte President of the Republic of South Africa 2000 2 SA 674 (CC) and Prinsloo v Van Der Linde 1997 3 SA 1012 (CC). Ackermann J said in Makwanyane (n 8 above) para 156 the following: ‘[W]e have moved from a past characterised by much which was arbitrary and unequal in the operation of the law to a present and a future in a constitutional state where state action must be such that it is capable of being analysed and justified rationally. The idea of the constitutional state presupposes a system whose operation can be rationally tested against or in terms of the law. Arbitrariness, by its very nature, is dissonant with these core concepts of our new constitutional order.’

25 Sec 9(5) contains a presumption that discrimination is unfair on any of the listed grounds unless proved to be fair.

26 See Albertyn & Goldblatt (n 14 above) 35 - 48.

27 Larbi-Odam v Member of the Executive Council for Education 1998 1 SA 745 (CC); and Khosa v Minister of Social Development; Mahlaule v Minister of Social Development 2004 6 SA 505 (CC).
underlies constitutional interpretation in South Africa, require that fundamental rights be extended to non-human animals and prohibit us from discriminating unfairly against them?

Some may be surprised by such a question, for indeed, human rights have often been understood to be simply those rights one has by virtue of being human. This understanding seems to be enshrined in the foundations of some of the most progressive international human rights instruments. The principle of equality is also often regarded as only applying to human beings and many writers simply fail to consider the application of this principle beyond the human species. Yet, the omission to do so does not constitute a justification: It can simply be an instance of moral blindness.

If we are to be consistent and reject one of the key moral failings of apartheid, we cannot simply be satisfied with a bare denial that non-human animals are entitled to equal treatment or fundamental rights. That would lead to dire consequences for a group of individuals without any consideration as to the justification for their exclusion from key legal protections. It would amount to a form of dogma that would use group characteristics (membership of a species) to discriminate against members of that group without any clear justification as to the relevance of that group characteristic. This would be to reproduce in our new order exactly the form of discriminatory treatment that the Constitution is designed to move away from. If differential treatment for non-human animals is to be justifiable, then this must be on the basis that they possess characteristics that do not entitle them to equal treatment.

The problem when we look scientifically at the world around us is that we find that many non-human animals possess many of the same characteristics that human beings have. Some are able to experience pain, many experience emotions, yet others have a rich cognitive life and yet others have strong social bonds. Many animals also exhibit the capacity for significant forms of agency and cognitive learning. These are the very characteristics that in the case of human beings entitle them to the protection of certain fundamental rights. And though many human beings possess other characteristics such as moral autonomy and self-consciousness, these latter

28 See J Griffin On human rights (2008) 13 who refers to this as the Enlightenment notion of human rights.
29 The International Covenant on Civil and Political Rights http://www2.ohchr.org/english/law/ccpr.htm, eg, recognises that the fundamental rights it proclaims derive from the 'inherent dignity of the human person'. The same is true of the International Covenant on Economic, Social and Cultural Rights http://www2.ohchr.org/english/law/cescr.htm.
30 Albertyn & Goldblatt, eg, do not consider the application of the equality clause to non-human animals. They are not alone and even a luminary philosopher such as Ronald Dworkin who places equality at the foundation of his political philosophy simply appears to ignore this question: See R Dworkin Sovereign virtue (2003) 1 - 7.
31 For the range and variety of mental states that may be attributed to other animals, see D DeGrazia Ticking animals seriously: Mental life and moral status (1994) 97 - 210.
32 DeGrazia (n 31 above) 158 - 165.
characteristics do not seem relevant to the provision of certain types of protections. The fact that a mentally ill individual may lack moral autonomy or self-consciousness does not mean that such an individual lacks the capacity to experience pain or emotions or social connections. Generally, we recognise that such mentally-ill individuals have certain rights: rights not to be subjected to pain, rights to express their emotions, rights to enable them to live lives that are able to express their social bonds. This must mean that having moral autonomy or self-consciousness is not necessary for having certain rights. If this be the case, and capacities to experience pleasure and pain, an emotional life, and/or social bonds provide the underlying grounds for possessing such rights, then it is unclear why non-human animals who have these characteristics are not equally entitled to such rights. The denial of such rights to non-human animals would only be justifiable if there were some good reason to do so. Yet, the only reason appears to be the intuition of some individuals that they are non-human. Yet, that comes down ultimately to an arbitrary preference for the human without any grounding in the actual nature of the beings in question.

If the Constitution were to include an arbitrary preference for the human without any sufficient justification, then it will simply reproduce the moral failings underlying apartheid in relation to a different group (those who are not human). This would fail to realise the Constitution’s promise as a transformative document, designed to place future South African society on a different and better moral foundation.

The Constitution must therefore be interpreted in such a way as to avoid this severe moral blunder. Currently, as it stands, the Constitution provides no express protections to non-human animals. The common law has traditionally classified animals as being forms of property who lack the

---

33 They are obviously relevant to some kinds of protections such as those relating to freedom of expression, association and religion.

34 This is a version of what may be termed ‘the argument from marginal cases’. The argument is found in a variety of literature on the subject: Examples include P Singer Animal liberation (1995) 16 - 20 and P Cavalieri The animal question (2001) 76.

35 These issues raise the question often posed in philosophy of the characteristics that ground ‘moral status’. T Metz ‘For the sake of friendship: Relationality and relationship as grounds for beneficence’ (2010) Theoria 54 - 76 suggests a theory – a form of ‘relational theory of moral status’ – that grounds moral status in relational properties that exist in the world rather than in certain intrinsic properties. It is not clear whether relational properties are fundamental and can ground a satisfactory view of moral status; even if they could, it is unclear that they do so in such a way as to include all and only humans. A full critique of this view lies beyond the scope of this paper.

36 This follows from the purposive and contextual approach adopted by the Constitutional Court. Metz, in his response to this article, (T Metz ‘Animal rights and the interpretation of the South African constitution’ in S Woolman & D Bilchitz (eds) Is this seat taken? Conversations at the Bar, the bench and the academy about the South African Constitution (2012) 209 claims that my interpretive approach fails to respect certain principles of ‘formal justice’ such as the ‘fit’ requirement that Dworkin elaborates. This means that interpretations need to be based upon political principles that make sense of the ‘recent history of their legal systems as a whole’ and are consonant with formal features of the system such as precedent. I shall raise two main points in response: First, the ‘fit’ requirement needs to be understood in the context of South African
capacity to have rights in their own right. Though some protections are given to animals through the Animal Protection Act, legal scholars have been divided as to whether these protections confer rights on animals. The *grundnorm* of interpretation argued for in this section means that the Constitution must be interpreted so as to ensure that no individual is subjected to differential (and prejudicial) treatment without any adequate justification. In light of this fundamental principle and, if the Constitution is not to be a fundamentally flawed and inconsistent document, it must be read in such a way so as not arbitrarily to discriminate against non-human animals. The next section seeks to employ the underlying *grundnorm* to explore the interpretive possibilities within the Constitution for granting animals the protections they deserve.

2 Interpretive possibilities in the Constitution for extending rights protection to non-human animals

2.1 Applying the Bill of Rights to non-human animals

The provisions of the Constitution that would be of greatest importance to non-human animals are contained within the Bill of Rights. Fundamental history and the Constitution. There has been a very short history of a just democratic order in South Africa and a long history of law rooted in injustice. Indeed, there are only 15 years of jurisprudence thus far by the Constitutional Court and no cases by that court on the question of the rights of non-human animals. In this light, the 'fit' requirement has to be understood differently to systems in which there is a long history of constitutional democracy and jurisprudence. In this context, it would in many ways be linked to the 'value' requirement but require a connection to the transformative purposes of the Constitution: A decision will 'fit' with the recent history of the legal system if it gives expression to the moral principles required to transform South African society from its apartheid past. Secondly, I would contend that the interpretive approach I have adopted does in fact meet the requirements of formal justice in South Africa: Indeed, it gives concrete expression to and is fully consistent with a key transformative ideal of the Constitution. Hence, I have sought to root my approach in South African history and the moral lessons we can draw from it that provided the catalyst for the new constitutional order. The test of an interpretation with 'integrity' is not whether the implications in question are accepted by everyone in society nor whether they reproduce the *status quo*: what is crucial is whether it faithfully constructs the Constitution in light of the underlying principles of the new Constitutional order itself. I have sought to show why a perhaps surprising conclusion for some about non-human animals flows from the deep purposes of the new constitutional order itself. In this sense, my interpretation seeks to exemplify the virtues of both 'fit' and 'value' as explicated by Dworkin. Indeed, he states in *Law's empire* (1986) 228 that '[w]hen a judge declares that a particular principle is instinct in law, he reports not a simple-minded claim about the motives of past statesmen, a claim a wise cynic can easily refute, but an interpretive proposal: that the principle both fits and justifies some complex part of legal practice, that it provides an attractive way to see, in the structure of that practice, the consistency of principle integrity requires'. This article seeks to argue that a recognition of animal rights and non-discrimination on grounds of species in the Constitution represents, in the context of South African history and constitutional practice, the realisation of 'the consistency of principle integrity requires'.


I have sought to canvass this debate in D Bilchitz 'Moving beyond arbitrariness: The legal personhood and dignity of non-human animals' (2009) 25 *SAJHR* 38 43 - 50.
rights provide protection for an individual’s most basic interests:39 If animals are excluded from having such protections, they become vulnerable to severe abuse. Immediately, the question that arises in this context is whether or not the Bill of Rights applies to non-human animals.

Section 8 directly addresses the application of the Bill of Rights. In relation to those who have duties imposed by the Bill of Rights, the section is clear: Section 8(1) provides that the Bill of Rights is binding on all organs of state; section 8(2) provides that the Bill of Rights is in fact binding upon natural and juristic persons to the extent it is applicable to them. Yet, in relation to the beneficiaries of the entitlements in the Bill of Rights, there is a noticeable gap in the text. Section 8(4) provides that ‘[a] juristic person is entitled to the rights in the Bill of Rights to the extent required by the nature of the rights and the nature of that juristic person’. Thus, the Bill of Rights expressly applies to juristic persons: Yet, it would be absurd to suggest that the fundamental rights contained therein were primarily designed to protect such entities as companies and non-governmental organisations. There is thus a clear omission in the application clause of the Bill of Rights, in that it does not define who, in general, are the beneficiaries of the rights contained therein.

Since there are no general provisions in this regard, it is necessary to interpret the Constitution to determine the beneficiaries of the rights contained therein. In doing so, it is necessary to consider this question in the context of each right. Certain rights are clearly demarcated as being restricted to certain beneficiaries: For instance, political rights are conferred only on citizens (section 19). The same is true of citizenship rights (section 20) and freedom of trade rights (section 22). Certain of the labour rights are restricted to employers, workers and trade unions respectively (sections 23(2) to 23(4)); children’s rights are only applicable to children (section 28), and the rights in section 35 are of application only to those who are arrested, detained and accused. The remaining rights in the Bill of Rights are phrased such that they benefit ‘everyone’: For instance, ‘everyone has the right to life’ (section 11). This phrasing of the rights replaced the use of the term ‘every person’ in the interim Constitution. Whilst Woolman attributes this change to the move towards more natural locutions in the final Constitution, he recognises that the extension of this term is ‘fuzzy’.40 This means that it is important to interpret what lies within the extension of this term. There is some case law that has sought to fill out who is included within the category of ‘everyone’. Non-citizens of South Africa have, for instance, been recognised as beneficiaries of certain rights in the Bill of Rights.41 There is no definitive decision concerning

---
39 For the relationship between fundamental interests and fundamental rights, see D Bilchitz Poverty and fundamental rights (2007) 47 - 74.
41 See (n 27 above).
whether a foetus is a beneficiary of the protections of the Bill of Rights, though the one existing case that addressed the question denied that a foetus would be entitled to such protections. In general, to whom then do the protections of the Bill of Rights apply? Could ‘everyone’ include non-human animals?

Using the fundamental grundnorm identified in the first section of this article suggests an answer to this question: We must avoid arbitrarily excluding anyone from the protections afforded by fundamental rights. As such, ‘everyone’ in the Bill of Rights should be interpreted in the context of each right and be inclusive of every individual with the fundamental interests necessary to be capable of benefiting from a particular right and needing its protection. Given that non-human animals have fundamental interests which require protection, this interpretation provides an opening to enable them to be beneficiaries of the entitlements afforded by the Bill of Rights.

It is important to clarify that not all rights will be of application to animals and some will only apply to a partial degree. This is an important point and, significantly, applies to human beings and other legal entities as well. Section 8(4) provides a useful illustration of this point in relation to the rights of a company which it provides must vary with the nature of the entity concerned: thus a company may need certain rights to privacy or to property, but it cannot really have a right not to be tortured or the right to vote. Though there is no similar provision applying to natural persons, it is clear that such a provision must be inferred: Whilst infants may have a right to freedom and security of the person, they cannot claim political rights which are meaningless to them. Similarly, rights to freedom of religion and to trade and occupation do not generally apply to them given their lack of capacity to exercise such rights. In the same way that infants do not have the right to vote or to trade, it would be meaningless for a chimpanzee to be granted such rights. However, the right to be free from all forms of violence, and the right not to be tortured are critically important both for infants and for chimpanzees. Similarly, many animals may also require certain rights to live in a habitat appropriate to their needs (a variant of the right to property), the right not to be deprived of the means of living (at least certain negative protections for socio-economic rights for wild animals), and the autonomy to live lives that are good in their own terms without undue human interference (certain freedom rights would thus be applicable to them). Thus, it is clear that

43 See, eg, Investigating Directorate: Serious Economic Offences v Hyundai Motor Distributors 2001 1 SA 545 (CC); First National Bank of SA Ltd t/a Wesbank v Commissioner, South African Revenue Service; First National Bank of SA Ltd t/a Wesbank v Minister of Finance 2002 4 SA 768 (CC).
44 There is simply no interest in these beings having the right to vote and consequently there is no right: ‘Nothing is blighted when a rabbit is deprived of the right to vote, or a worm of the free exercise of religion’ (M Nussbaum Frontiers of justice (2006) 361).
applying the Bill of Rights to animals does not lead to absurdity but only to
the affording of protections for their most fundamental interests, many of
which they share with humans.

Though a reading of the term ‘everyone’ that includes animals does not
strike me as overly forced, there are certain objections that must be
responded to, if it is to be entirely persuasive. The first problem relates to the
fact that, traditionally, the common law has recognised that only ‘natural’
and ‘legal persons’ are entitled to have rights. Traditionally, it has been
that all and only human beings have been recognised as being ‘natural
persons’ with non-human animals being classified as ‘things’. It would
thus seem to go against established doctrine in South African common
law to allow animals in the Constitution to claim the protection of
fundamental rights where they have been classified simply as ‘things’.

The difficulty with this argument is that it fails to recognise the
transformative effect the South African Constitution is meant to have on
the common law. The Constitution was not meant simply to replicate the
existing law but to change it to represent the values of a new era. Importantly, section 39(2) states that ‘[w]hen interpreting any legislation,
and when developing the common law or customary law, every court,
tribunal or forum must promote the spirit, purport and objects of the Bill of
Rights’. It is thus a duty of judicial officers to develop the common law in
light of the spirit, purport and objects of the Bill of Rights. As was argued
in section 1 of this article, one of the key principles giving content to the
‘spirit, purport and objects of the Bill of Rights’ is a commitment to
equality and non-arbitrariness. The common law must therefore be
developed in light of that commitment: If the classification of animals in
the past as ‘things’, with no rights of their own was arbitrary, then that needs
to be rectified in the new constitutional order. The Constitution can thus
be seen to require the development of the common law to include non-
human animals within its ambit. This could be achieved in legal doctrine
in several ways.

First, the very vagueness of the term ‘everyone’ enables rights to
be extended to animals without necessarily finding that they should be
classified as ‘persons’. This approach would simply mean that the status of
animals will have to change in the common law without recognising what
their new status would be. It may be that something in between ‘things’ and
‘persons’ would be appropriate: A category of ‘beings’ could be developed,
for instance, which includes those who are capable of having rights but not

46 Sinclair (n 37 above) 4 - 5.
47 I have sought to argue against the rationality of this classification (even without the
assistance of the Constitution) and in light of the Animal Protection Act in Bilchitz (n 38 above) 43 - 50 and 65 - 68. For some judicial support for this line of argument, see
Cameron J’s judgment in National Council of Societies for the Prevention of Cruelty to Animals v Peter Openshaw (462/07) 2008 ZASCA 78 RSA (NSPCA).
exercising duties. Indeed, recognising such an intermediate category may be useful in the new constitutional order as it would enable us better to recognise the differences between human beings with differing capacities: Those who are capable of exercising duties would, for instance, be entitled to all the rights in the Constitution, whilst those only capable of having rights and no duties would be entitled to a more limited selection depending on their capacities. The current category of the ‘person’ includes both categories of individual without recognising the differences between them.48

However, if judicial officers prefer for the law to retain the binary classification of ‘persons’ and ‘things’, then it may be necessary to use the ethos underlying the new constitutional order to have animals classified as ‘persons’. Indeed, once we recognise that small children and the mentally-ill are classified as persons, it becomes clear that those capable of only exercising rights and not duties have already been included within this category. As such, there is no clear conceptual impediment to including animals with similar capacities within the category of being a ‘natural person’ and this would be in line with the development of the common law in light of the new constitutional order.49

There is another approach that would be viable in terms of our constitutional framework. If an artificial legal fiction such as a company can be entitled to protections under the Bill of Rights, then a fortiori it appears that sentient beings with fundamental interests that can affect their very capacity to enjoy decent lives, should be entitled to such rights. Indeed, reasoning such as this has led to the suggestion by some authors that animals be classified as a particular type of ‘juristic person’ rather than a ‘natural person’.50 Since the law may confer legal personality on any entity that it wishes, and there is a strong claim on the part of animals to be protected by law in their own right, this may offer a solution that would extend legal personality to animals with no shift in the traditional meaning of the ‘natural person’ in our law. Whilst this approach is certainly possible, this taxonomy does not, however, strike me as ideal. It makes more sense to place similar entities in the same category: since humans and animals are both natural entities with intrinsic value in their lives and with their own welfare, they belong in the same category and should be considered ‘natural persons’. They are unlike companies where legal personality is in fact a fiction.

I have thus sought to suggest a number of possibilities for the required doctrinal shift in the law in the classification of animals: A decision between these various approaches, however, will depend on the

48 See the critique of the current concept of the person in Bilchitz (n 38 above).
49 As above.
50 See, eg, D Favre ‘A new property status for animals’ in C Sunstein & M Nussbaum (n 18 above) 244 - 245.
temperament of judges for change and the degree to which they are prepared to shift the traditional categories of the common law.\textsuperscript{51}

It could be objected that the following textual provision in the Constitution suggests difficulties for the view that the term ‘everyone’ in the Constitution could include non-human animals. For section 7(1) of the Constitution provides that the Bill of Rights ‘is a cornerstone of democracy in South Africa. It enshrines the rights of all people in our country and affirms the democratic values of human dignity, equality and freedom.’ A plain language reading of this provision, it could be argued, suggests that it is only ‘people’ who are entitled to the protections contained in the Bill of Rights.

In response, it is important to recognise the significant impact the ‘purposive’ and ‘contextual’ interpretive method can have on the interpretation of a provision in the Constitution.\textsuperscript{52} The seemingly plain reading of a provision is not determinative and may in fact be modified by such an approach. Section 7(1) itself expresses a commitment to the values of human dignity, equality and freedom. I will consider some of these values further below but, importantly, this provision itself and these values can be seen at least partially to be an expression of the \textit{grundnorm} which requires that no individual be subjected to arbitrary treatment. As has been argued, this would require the inclusion of non-human animals in the category of rights-bearers. This means that the statement ‘enshrining the rights of all people’ in our country must be interpreted to allow for this possibility. Two interpretive strategies can be adopted: First, if ‘people’ is taken to refer to human beings alone, it can be argued that this provision does not seek to exclude consideration of whether other beings are also entitled to the rights in the Bill of Rights. The provision in question is an introductory provision and can simply be understood to indicate the shift the Constitution has brought about from a past where the rights of all people were not adequately protected to a situation in which they are afforded such protections. The term ‘everyone’ in the Bill of Rights would thus apply to a broader class of individuals than simply ‘people’ and could thus include animals. Secondly, a more radical interpretation would follow the reasoning outlined above and contend that the extension of the term

\textsuperscript{51} Given the judicial culture of minimalism in South Africa, I have sought to outline the possibilities such that judges could choose an approach that they regard as involving the least interference with the common law position. It must though be borne in mind that the Constitution does mandate being prepared to transform the common law.

\textsuperscript{52} For a famous example of courts going beyond what appeared to be the plain meaning of a text, see \textit{S v Mhlungu} 1995 3 SA 867 (CC). This case has generated much academic debate: See, eg, A Fagan ‘In defence of the obvious – ordinary meaning and the identification of constitutional rules’ (1996) 12 \textit{SAJHR} 545 - 570; E Fagan ‘The longest erratum note in history’ (1996) 12 \textit{SAJHR} 79 - 89. D Davis ‘The twist of language and the two Fagans: Please sir may I have more literalism’ (1996) 12 \textit{SAJHR} 504 - 512; E Fagan ‘The ordinary meaning of language – a response to Professor Davis’ (1997) 13 \textit{SAJHR} 174 - 177; D Davis ‘Of closure, the death of ideology and academic sand castles – a reply to Dr Fagan’ (1997) 13 \textit{SAJHR} 178 - 180.
‘people’ depends upon who is included within the common law categories of ‘natural’ or ‘juristic’ persons. If these categories are extended to include animals, then the term ‘people’ could include them as well and therefore does not pose any obstacle to extending rights to them.53

2.2 Dignity

One of the possible stumbling blocks in the way of an interpretation that would apply the rights in the Bill of Rights to non-human animals could be the way in which the value of dignity is expressed in the Constitution. Section 1(a) speaks about the foundational value of ‘human dignity’ and the value of ‘human dignity’ is affirmed in section 7 of the Bill of Rights. Section 10 states that ‘[e]veryone has inherent dignity and the right to have their dignity respected and protected’. Though the actual content of the right does not refer specifically to human beings, the heading of the right in the Constitution is ‘human dignity’. This would seem to imply that only human beings are covered by the right in question. Finally, section 36(1) provides that rights may be limited only by a law of general application in circumstances where this would be reasonable and justifiable within an open and democratic society based on human dignity, equality and freedom.

In all instances where dignity is mentioned in the Constitution, it appears to be confined to human beings. This could be said to suggest that the focus of fundamental rights provisions that centrally protect the dignity of individuals is upon human beings. Importantly, the usage of the term human dignity in the South African Constitution is consistent with international human rights instruments.54 The historical context of these legal documents, however, provides once again a clearer understanding of the origins and purpose of the provisions relating to ‘human dignity’.

At the international level, the Universal Declaration of Human Rights arose in light of the holocaust in which there had been a massive denial of the equal worth of all human beings. The Nazi regime had sought to distinguish between different classes of humans, with only Aryans having equal worth and all others being inferior. Such denials of equal dignity for all led to mass murder in the case of Jews, gypsies and certain other groups whilst justifying the enslavement of groups such as Slavs. It was against the backdrop of this history that the human rights instruments adopted after the World War II sought to assert the fundamental equal worth of all human beings in order to prevent such events from happening in future and

53 This latter proposal does not accord as well with ordinary language usage, though such usage is not determinative of meaning in the Constitution.
as a foundation for the fundamental rights that they proclaimed. The post-World War II German Constitution also consequently included human dignity in its very first article and provided that it shall be ‘inviolable’ and that it is the duty of all state authority to respect and protect it.

South African history also saw the denial of the equal dignity of human beings, with the whole structure of society and governance being based upon the superiority of whites and the inferiority of black people. This in turn led to the massive denial of rights in the case of black people and a range of practices – such as the ‘pass laws’ – that imposed humiliation and insult upon them. The assertion of equal dignity in the Constitution represents a repudiation of this history and a strong commitment to ensure that no human being may be subject to such reprehensible treatment in future. ‘[J]ust as the Germans have promised not to shovel people into stoves, so too have South Africans promised never again to treat people like cattle to be packed off to Bantustans or to be slaughtered in the middle of the night.’ The Constitutional Court has in several cases confirmed this understanding of the meaning and function of dignity in our constitutional order. Justice O’Regan, for instance, gives clear expression to this idea when she writes that ‘[t]he Constitution asserts dignity to contradict our past in which human dignity for black South Africans was routinely and cruelly denied. It asserts it to inform the future, to invest in our democracy respect for the intrinsic worth of all human beings.’

In the light of South African history, such a statement is clearly poignant and of crucial importance. It perhaps represents a further central principle of the Constitution that human beings should never be treated again in a manner that fails to respect their dignity. Yet, as we have seen, that very history mandates a strong ‘equality principle’ in the interpretation of our Constitution that involves not arbitrarily subjecting any individual to prejudicial treatment. If we are not to have a normative conflict at the heart of our Constitution, these two principles must be reconciled with one another. It will not do simply to ban arbitrary discrimination amongst human beings for, as has been argued in section 1 above, that would undermine the very insights of the equality principle.

55 As Woolman has put it, ““dignity” is the flip-side of “never again”:’ See ‘Dignity’ in S Woolman et al Constitutional law of South Africa (2006) 36-4. For this reason, he claims that it is no accident that it plays a central role in German jurisprudence.
57 Woolman (n 55 above) 36-4.
58 Dawood, Shalabi and Thomas v Minister of Home Affairs 2000 3 SA 936 (CC) para 35.
and allow for prejudicial treatment simply on the basis of the group (of *homo sapiens*) to which an individual belongs.\(^{59}\)

Another option, however, suggests itself. This involves understanding that whereas the Constitution does refer specifically to ‘human dignity’, this does not seek to deny the possibility that other types of beings may also have worth in their lives and, thus, consequently, be entitled to the protection of certain fundamental rights provisions. The explicit textual focus on *human* dignity can be explained through understanding the proximal historical context in which the Constitution was adopted, namely, to repudiate the massive denials of human dignity under apartheid and to signal that, in future, all human beings would be treated as being of equal value.\(^{60}\) Against this historical backdrop, it does not appear that the focus on ‘human dignity’ in the Constitution was designed to take a position on the question whether *only* human beings have worth or specifically to exclude non-human animals from being entitled to the protections in the Bill of Rights. It was meant simply to assert the fact that *all* humans have dignity without necessarily implying anything in relation to other creatures. Whilst the specific provisions relating to human dignity will not apply to non-human animals, understanding the historical context of these provision opens up the space for an interpretation of the Constitution that would acknowledge the worth of other creatures and enable them to be provided with the protections afforded by the Bill of Rights.\(^{61}\)

The importance of not confining assumptions of worth to human beings and providing some form of constitutional protection for non-human animals can be illustrated by a case that arose in Germany. As has been mentioned, the German Constitution incorporates a strong emphasis on human dignity and the Federal Constitutional Court has provided a rich understanding of this notion.\(^{62}\) The rights in the German Constitution had traditionally been recognised as extending only to human beings. Litigation arose before the courts on whether a Sunni Muslim butcher was entitled to slit the throat of an animal which, according to his religious beliefs, had to occur without stunning it beforehand. There were many Muslim religious authorities that allowed for pre-stunning, yet this particular sect believed it to

---

\(^{59}\) Indeed, C McConnachie ‘Human dignity is an oxymoron: An argument for a non-speciesist understanding of dignity’ (unpublished paper on file with author) argues that the very notion of human dignity contains an internal contradiction and the underpinning ideas behind this concept require extension of our ethical duties beyond the human species.

\(^{60}\) Woolman (n 55 above) 36-2 - 36-4.

\(^{61}\) In Bilchitz (n 38 above) 50 - 65, I provide a fuller account of the problems with confining a notion of ‘dignity’ or ‘worth’ only to human beings. Using the recent work of Nussbaum (n 44 above), I attempt to provide an account of how animals may also be said to have ‘dignity’. The focus of this article is, however, on the way in which historical context can help us ground an interpretation of the Constitution that does not arbitrarily discriminate against non-human animals.

\(^{62}\) See Woolman (n 55 above) 36-1.
Does transformative constitutionalism require the recognition of animal rights?

be impermissible. A constitutional argument was raised by the Sunni Muslims in question claiming that their right to slaughter without pre-stunning was an incident of their right to freedom of religion. This argument succeeded in the Federal Constitutional Court largely as a result of the fact that there was no countervailing constitutional interest on the part of the animals that the court could take into account. The rights of the human beings in question were absolute in relation to the animals concerned as the animals had no interests that were protected at the constitutional level and that could be balanced by the court against those of the human beings in question. If we are adequately to capture the complexity of the moral evaluations to be undertaken in a case such as this, the interests of the animals must at least be capable of holding some weight in a court’s deliberations. Even if ultimately the human interests in such an instance would outweigh those of the animals, the problem lies in the fact that without some form of constitutional protection, there was no balancing at all to be done in legal terms. The difficult moral evaluation in question could not be articulated in constitutional terms. As a result of this case, the German legislature introduced an amendment to the Constitution which recognised the interests of animals and the environment.

Whilst the wording of the clause in question is far from optimal, it allows for animal interests to be protected at the constitutional level and thus provides the scope for legal argumentation that could, in certain cases, limit human rights on the basis of animal interests.

This case illustrates the importance of recognising that animals have some worth and interests which must be taken into account at the constitutional level. Failure to do so leaves animals at the mercy of humans who can often attempt to defend their actions on grounds of fundamental rights. Without any countervailing constitutional interests on the part of animals, invariably, the humans will triumph simply because they belong to the group of homo sapiens. This would preclude a proportionality inquiry into which interests should take precedence and always places human interests above those of animals even where the substantive considerations at issue are in favour of the animals: To do so, would thus clearly violate the grundnorm proscribing prejudicial treatment against

---

63 The account of this issue is based on M Nattrass ‘… Und die tiere: Constitutional protection for Germany’s animals’ (2004) Animal Law 283.
64 See Grundgesetz art 4.
65 Nattrass (n 63 above) 301.
66 See Grundgesetz article 20a: ‘[M]indful also of its responsibility toward future generations, the state shall protect the natural foundations of life and animals by legislation and, in accordance with law and justice, by executive and judicial action, all within the framework of the constitutional order’ (translation by Sebastian Seedorf). The recognition, however, is not adequate as it fails to grant rights to animals and simply places animal interests as a value and goal to be taken into account by the state – something that would nevertheless be progressive in South Africa where animal interests are often disregarded. See Nattrass (n 63 above) for a discussion as to its history and implications, some of which are relevant to South Africa.
individuals without any adequate justification. The South African Constitution must thus be interpreted to avoid such a result and allow for the worth of other creatures and their entitlement to certain fundamental rights to be recognised.\textsuperscript{67}

2.3 Equality

Whilst the notion of human dignity may not provide an obstacle to the recognition of animal rights, it does not actively advance the case for such recognition. The value and right to equality can provide such a positive case. I have in section 1 already outlined ‘the equality principle’ and the case flowing from this for the recognition of animal rights. It is important here to recognise that this case can be supported by some of the features of the rich jurisprudence that the Constitutional Court has developed around the equality right.

The first important feature of this jurisprudence is the requirement of justification. Section 9(1) – which provides that ‘everyone is equal before

\textsuperscript{67} Metz in Woolman & Bilchitz (n 36 above) 214ff, in his response to this article, has objected to my attempt to interpret the Constitution in this way through arguing that such reasoning might violate obligations of compensatory justice that we have to black Africans. I am in agreement that compensatory justice is important in South Africa and that we need to take more account of African values and make amends for the wrongs perpetrated in the past in relation to African culture. I take issue though with the argument by Metz that compensatory forms of justice in South Africa require excluding animals from being recognised as having worth or being entitled to fundamental rights protections. First, it is important to point out that it is not only African culture that has often ignored animal interests and arbitrarily failed to take them into account. Discrimination against animal interests is widespread in many cultures and can be likened to other phenomena such as patriarchy and heterosexism. If we are to respect a culture, this does not mean adopting a static view of that culture and accepting all features of that culture no matter how damaging to women, gay people or other creatures. African culture must definitively be promoted but the elements of it that are patriarchal, heterosexist and I would add ‘speciesist’ need to be shifted in light of the new constitutional order. This is something the Constitutional Court has recognised in the context of women’s rights in the case of \textit{Bhe v Khayelitsha Magistrate} 2005 1 SA 580 (CC), where it recognised the need for customary law to develop in a dynamic manner and to do so in a way that avoids entrenching historical patterns of disadvantage in relation to women. Secondly, and perhaps most importantly, we should reject the moral principle that we can compensate for wrongs done to a particular group (black people) by perpetrating on a massive scale wrongs against another group (non-human animals). If we assume, eg, that patriarchy has historically been a part of African culture, Metz’s reasoning would lead us to the absurd conclusion that to compensate black people for the wrongs done to them, we must perpetuate the wrongs that have been done in their culture to women. Indeed, it could convincingly be argued that groups such as women, gays and animals have claims for compensatory justice as well in light of wrongs done in the past: Why does Metz only see compensatory justice as applying to black people? I have sought instead to argue that transformative constitutionalism requires that we reject the particular moral wrongs of apartheid not only in relation to race but in relation to all groups. It is certainly not clear to me that African culture (or any other culture) will not survive such a shift. Finally, as Metz himself acknowledges, there is some dispute concerning the place of animals within African culture (see his footnote 10) and thus, in light of developing scholarship in this area, it certainly cannot be concluded that the shift I am arguing for is clearly out of step with African moral theory (which itself is a very broad domain).
Does transformative constitutionalism require the recognition of animal rights?

law’ – has been interpreted to require that when any differentiation between individuals or groups occurs in law, there must be a rational connection between the differentiation and a legitimate government purpose. Thus, this requires consideration of the purpose of a provision, whether it is legitimate and whether the differentiation furthers the purpose. For animals this can itself provide some protection for, in many cases, there is very little thought given to differentiations in treatment that occur in relation to non-human animals and whether they are rationally connected to a legitimate purpose or not.

The Court’s jurisprudence also contains another key feature of importance in this context in its analysis of what constitutes unfair discrimination in section 9(3). The Court states that ‘in the final analysis it is the impact of the discrimination on the complainant that is the determining factor regarding the unfairness of the discrimination’

In determining impact, the Constitutional Court outlines three important factors: first, the position of the complainant in society, and whether they have suffered as a result of past patterns of disadvantage; secondly, the nature of the provision or power and the purpose sought to be achieved; and finally, the extent to which the discrimination affects the rights and interests of the complainant and whether it has led to an impairment of their fundamental human dignity or constitutes an impairment of a comparably serious nature.

Importantly, discrimination against non-human animals often has a dire impact on them. Their interests are routinely ignored and they are treated as mere inputs into an industrial process. The factors outlined by the Constitutional Court are also useful in assessing impact in relation to animals. Indeed, they have been treated in detrimental ways since time immemorial. Legal regulation and practices have historically been designed largely around the human benefits to be obtained from exploiting them without a consideration of their own interests and needs. Animals are also profoundly vulnerable and human beings have the potential to determine whether they are able to lead decent lives or in fact live at all.

68 Harksen v Lane 1998 1 SA 300 (CC) para 42.
69 Harksen v Lane (n 68 above) para 49.
70 Harksen v Lane (n 68 above) para 50.
71 The South African Nobel laureate for literature, JM Coetzee, accurately captured the oppressive use of animals in modern factory farming when he compared it to the destruction of humans in the holocaust and stated the following: ‘What a terrible crime to treat human beings like cattle – if we had only known beforehand. But our cry should more accurately have been: what a terrible crime to treat human beings like units in an industrial process. And that cry should have had a postscript: What a terrible crime – come to think of it, a crime against nature – to treat any living being like a unit in an industrial process.’ This article ‘Exposing the beast: Factory farming must be called to the slaughterhouse’ can be found at http://www.smh.com.au/news/opinion/exposing-the-beast-factory-farming-must-be-called-to-the-slaughterhouse/2007/02/21/1171733846249.html.
The second factor outlined by the Court is similar to the justification requirement of any law but the third factor can also have application to non-human animals. Importantly, the court does not confine a consideration of the effect of the discrimination to the impact on human dignity alone, but recognises the possibility that there may be ‘an impairment of a comparably serious nature’. This aspect of the Constitutional Court’s jurisprudence widens the scope of what may constitute unfair discrimination. An interesting implication though is that this would seem to open the possibility that other forms of life may also have worth and be harmed in a manner comparable to a detrimental effect on human dignity. Indeed, it would seem clear that many forms of treatment accorded to animals grant little or no consideration to their interests and so treat them in a manner not consonant with their worth. These forms of treatment would appear to be prime candidates for impacting upon animal interests in a comparably serious manner to an impairment of human dignity.

The Constitutional Court’s jurisprudence on equality thus contains within it the resources to extend the protection of the equality and non-discrimination clauses of the Bill of Rights to non-human animals. An important challenge, however, needs to be considered here. It is indeed conceivable that many people will be prepared to acknowledge my claims that animals have worth, and are entitled to certain protections. Yet, such people may contend that animals are not entitled to exactly the same protections as human beings as their interests do not carry equal moral weight to those of human beings. Such a line of reasoning may be supported by the fact that many people have in their mind a ‘Great Chain of Being’, recognising that some non-human animals have more ‘worth’ than others with humans at the top of the chain. The problem, such an individual may claim, is not with recognising the worth of animals but recognising their equal worth and consequently our being required to grant their interests equal moral weight. Are there any good grounds to resist such an argument?

In responding to this argument, it is important to revert to the grundnorm requiring us not to treat individuals differently without adequate reason. Importantly, this norm places the presumption firmly in favour of equal moral weight being given to the interests of differing individuals. Those wishing to argue for lesser weight to be granted to animal interests must do more than claim that this is what we generally have thought. For, indeed, the history of claims relating to differential moral weight being attached to interests has often led to grave forms of injustice, including those under both the Nazi and apartheid systems. This history draws our attention to the way in which claims of differential moral status have been abused and so should make us extremely cautious before accepting that individuals of different groups are of differential worth unless there is a convincing case to this effect.
In relation to non-human animals, specifically, it is clear that we have certain strong ingrained prejudices in favour of humans which again should make us particularly wary against concluding that the interests of such creatures should be granted lesser weight. Inevitably, judgments that impact upon laws are always going to be made by human beings in relation to non-human animals: There is thus an inescapable likelihood of bias in favour of the human. Moreover, many humans have a strong self-interest in relegating animals to a second-class moral status which would enable human interests easily to outweigh animal interests. Many of these interests are very close to human beings and include the eating of meat, the development of medicines, clothing and entertainment. As a result of these interests, over time, human beings have developed a range of strategies to render them less sensitive to animal suffering. Intellectual honesty requires that we recognise already that there will be a strong likelihood of bias against animals in any decisions relating to them. These reasonable apprehensions of bias on our part reinforce the view that the presumption in this context needs to be firmly in favour of equal moral weight being attached to animal interests unless there are good reasons to depart from this. This presumption does not mean that animal interests would always win out: It simply means that in considering a law or practice that infringes on an animal’s rights, they must be given fair and equal moral weight in any constitutional evaluation.

It is important to point out, that a commitment to equal moral weight being attributed to animal interests does not entail identical treatment. The

---

72 Metz in Woolman & Bilchitz (n 36 above) 217 gives several examples with which he seeks to support the unequal moral status of humans and animals. The one example involves two humans (A and B) who are starving along with a pig. Metz claims that it would be permissible to shoot the pig to feed A and B, but it would not be permissible (indeed, it would be gravely disrespectful) for A to shoot B to feed himself and the pig. This example, suggests Metz, is best explained by the unequal moral status of the pig and the two humans. Even if we agree with Metz’s description of what should happen in this example, his conclusion does not follow. Consider an example where there are three humans, X, Y and Z who are all about to starve and are running out of rations. X and Y are roughly 25 years old and Z is 90 years old. One of the humans has to be allowed to die if the others are to survive. Most people it would seem would judge it better for Z to be killed than either X or Y on account of his age. Does this mean that older people have less moral status than younger ones?

73 Metz in Woolman & Bilchitz (n 36 above) 218 - 9, in his response to this article, clearly demonstrates the dangers of failing to grant equal moral weight to animal interests. He suggests that to expend any resources to enforce even the most minimal protections for animal interests could constitute ‘disrespect’ for Africans where so many are suffering. I have already provided a response to the compensatory justice argument: However, it is important to recognise that ‘disrespect’ could only be understood if one thinks that animals are not entitled to equal consideration in distributive decisions. Moreover, Metz’s point seems to suggest that no resources should be dedicated to protecting animal interests, thus, even failing to grant their interests minimal weight. The lack of equal concern thus translates into a lack of any form of concern at all from the point of view of public institutions. The simple point is that if animals count morally (and I have provided reasons for thinking they should count equally), some resources at least will have to be distributed towards defending their interests (if not actively advancing them). The alternative is arbitrarily to prejudice the interests of non-human animals, something proscribed by the grundnorm underlying our Constitution.
differential capacities and capabilities of different species mean that, at times, they may be subject to different treatment which is fully justified. This will not be a violation of the principles of equality but rather an expression of them. Thus, treatment must be related to the capacities of a creature: It is no violation of the worth or interests of a dog to deny it the right to vote; it would be a violation of those interests if individuals were not prevented from kicking it and so respecting its right to bodily integrity. Equal treatment of dolphins (and their rights to environment, property or housing) would require we protect their aquatic environment while in relation to elephants would mean we ensure they have sufficient space to roam. Often, what will be required of us in relation to wild animals at least is simply to desist from harming them rather than actively interfering in their lives. These examples simply seek to show that non-discrimination on grounds of species does not require the ‘same’ treatment in all cases: the treatment will depend on the nature of the interests in question.

This section has sought to develop a number of interpretive possibilities in the South African Constitution in light of the fundamental ‘equality principle’ that would support the extension of some of the rights therein to non-human animals. It should be clear, however, that there may be circumstances that arise of conflict between animals of different kinds and between animals and humans. The grund-norm of equality requires in such cases that there be a justification for any differential treatment that limits a right and that such a justification meet certain conditions. The next section seeks to analyse the manner in which the Constitution can recognise limitations on animal rights.

3 Transformative constitutionalism and the limitation of animal rights

If, as has been argued, animals are entitled to the rights in the Bill of Rights, then they will also be subject to its provisions concerning limitation. This would indeed follow from the ‘equality principle’ which would naturally

---

74 See the discussion in Nussbaum (n 44 above) 372 - 380.
75 This discussion clearly has not sought to be exhaustive but has dealt with more general values and rights in the Constitution that provide certain of the leading arguments for such an extension. A notable omission is the lack of discussion of sec 24 which contains an environmental right. Though animals could be included within the notion of the ‘environment’, the protections afforded by this section fail to distinguish between animals and other elements of the environment such as plants, stones or creatures to be treated as ‘things’ without interests of their own. The argument of this paper is that animals do in fact have their own welfare and their own interests in light of which it makes sense to recognise the principle that they should not be subject to prejudicial treatment without good reason. The grundnorm could perhaps be used to provide an interpretation of the environmental right that grants certain protection to animals. However, the approach I adopt seeks to go further and recognise the potential for many rights in the Constitution to have application beyond the human species. I have thus sought to concentrate on provisions that would enable me to make this wider case.
Does transformative constitutionalism require the recognition of animal rights?

require that animal rights are subject to the same principles concerning limitation as human rights. Importantly, though, there are wide-ranging human practices that currently infringe the rights of animals. Even though the 'equality grundnorm' underlying our new constitutional order may argue in favour of the abolition of such practices, it is unlikely that they will cease any time soon or that human beings in the South African polity are ready to desist from such common habits as eating meat. If the new constitutional order does not, in some way, require a concrete shift in these practices, then it is unclear what gains will in fact be achieved by recognising animal rights. Thus, it is of particular importance in this regard to articulate a conceptual framework concerning limitations through which the recognition of animal rights in the Constitution may have an impact upon current abuses without going too far beyond what the majority of South Africans are prepared to tolerate.

It is important to recognise that, in light of prevailing moral attitudes, there are two kinds of limitations that may be identified in the context of animal rights. The first kind of limitation would be one that would be akin to those recognised in the case of human beings: Such a limitation would flow from the differing normative claims and conflicts that arise in connection with rights and the requirement to balance competing claims in terms of the limitations clause. Such limitations would be fully justifiable in terms of the grundnorm underlying our new order and simply extend the same logic applicable to humans to other creatures.

A second kind of limitation, however, would arise not as a result of the requirements of balancing but as a result of the general inter-subjective acceptance by the majority of South Africans of a particular limitation on animal rights. The general inter-subjective acceptance of such a limitation does not entail that such a limitation is objectively justifiable. Nevertheless, even if it is not objectively justifiable, prior to the widespread recognition that this is so, it will be necessary for the maintenance of respect for the constitutional order to make provision for such widespread inter-subjective acceptance without fatally compromising the recognition of animal rights. Such an acceptance may not be ultimately consonant with the equality grundnorm, but be necessary to hold out the hope of a society that does ultimately fully instantiate the implications of such a norm. The following discussion shall consider two different conceptual features of the South African Constitution that can provide assistance in determining the limitations on animal rights in light of the considerations discussed. First, I shall consider the notion of 'progressive realisation' and, secondly, I shall consider certain elements of the general limitations clause that can assist in this regard.

3.1 ‘Progressive realisation’ and animal rights

Certain of the rights in the South African Constitution contain what is often referred to as ‘internal limitations’. This means that the rights can be limited by reference to the factors contained in the very provision giving expression to the right concerned without initially having to justify that limitation in terms of the general limitation clause. In relation to the socio-economic rights in the Constitution, for instance, this internal limitation clause requires the state to take ‘reasonable legislative and other measures, within its available resources to achieve the progressive realisation of these rights’.77 The Constitutional Court has given meaning to each of the qualifications contained in this clause: For present purposes, I wish to focus on the notion of progressive realisation.

This concept was introduced initially in the International Covenant on Economic, Social and Cultural Rights and has been developed by the UN General Committee on Economic, Social and Cultural Rights.78 This notion involves recognising that the full realisation of these rights may not be possible at a particular point in time: Yet, it nevertheless imposes an ‘obligation to move as expeditiously and effectively as possible towards that goal’.79 It also requires that deliberately retrogressive measures generally not be taken and that they be carefully justified if they become necessary.80 This obligation has also been linked to a minimum core obligation to ensure the satisfaction of minimum essential levels of each right.81 Thus, in the context of socio-economic rights, it may not be possible to provide everyone with a decent house immediately, but the government may nevertheless be obliged to ensure that individuals at least have shelter from the elements in the interim.82 The notion of progressive realisation was thus developed in the context of a scarcity of resources and the recognition that fully realising these rights will only be possible over time. It was designed also to ensure that such rights remain meaningful whilst recognising the pragmatic difficulties of ensuring their full and immediate implementation.

The purposes and content given to the notion of ‘progressive realisation’ suggest that it may be an important interpretive tool to use in developing South African jurisprudence surrounding animal rights and understanding the limitations thereof. Importantly, there will be some differences in the way the concept is applied in this context. In relation to

77 Sec 27(2), eg, contains this internal limitation.
78 See art 2 of the Covenant (n 29 above). It is also included in secs 26(2) and 27(2) of the South African Constitution.
80 As above.
81 General Comment 3 (n 79 above) para 10.
82 See Bilchitz (n 39 above) 193 - 194.
animal rights, for instance, what is at stake often in relation to wild animals is simply our non-interference with their natural environment and allowing them to conduct their own lives. Negative obligations are thus paramount though there may be some limited forms of positive involvement as well. In the context of domestic animals and animals in captivity, positive obligations assume greater prominence in that we are often required actively to provide domestic animals with food and shelter. Negative obligations that involve not subjecting an animal to cruel treatment nevertheless remain of importance. Though, the notion of ‘progressive realisation’ was developed in general in relation to positive obligations flowing from socio-economic rights, in the context of animal rights, I would argue, it can usefully be applied both in relation to the negative and positive duties we owe to animals. Courts would need to decide that the notion be of application beyond the express context of socio-economic rights (even without clear textual pointers). Doing so could have a number of advantages.

At present, it does not seem that South African society is ready to accept the full realisation of animal rights: Nevertheless, it may be possible to achieve some form of ‘partial realisation’ thereof. As has been seen, the ‘equality principle’ mandates that we give attention to ensuring that such rights are not simply arbitrarily curtailed. This partial realisation could involve the recognition of a minimum core of rights which could be developed to provide an understanding of what each animal is entitled to as a matter of priority. It is unlikely that the right to life would be part of such a core (given current societal expectations concerning meat-eating) but other rights of great importance for creatures could be seen to be part of that core: Indeed the right to bodily integrity and rights to sufficient food and water (when in the care of humans) would immediately seem to be candidates for such a minimum. Deliberately retrogressive measures against animals could be banned and a strict justification required for any such measures to be passed. The government could also be required to develop a plan progressively to extend and develop the rights animals have over time and to educate individuals to have a more respectful attitude towards them.

The concept of progressive realisation thus could be useful in providing a sense in which certain animal rights are protected now with the idea that their entitlements will be expanded over time. It also contains an in-built flexibility which may be useful in this context though also a subject of critique. For the concept does little in itself to provide the basis upon which we can distinguish between the rights (or features thereof) that will be partially

83 Nussbaum (n 44 above) 373 - 380 recognises that negative obligations may assume a greater importance in relation to animals but contends that there are still significant positive obligations upon humans.

84 This core could, eg, include rights that initially protect five freedoms developed by the UK Farm Animal Welfare Council: freedom from thirst and hunger, freedom from discomfort, freedom from pain, injury and disease, freedom to express normal behaviour, freedom from fear and distress. See http://www.fawc.org.uk/freedoms.htm.
realised now and those that will be subject to progressive realisation: The only helpful notion here is the idea of a core set of rights that are particularly urgent. Even then, the basis upon which such a core is determined will also need to be considered: It would seem that even here, pragmatic factors may come into a determination of this core.85 Without being developed further, the concept thus seems to have a number of inherent dangers: It is vague conceptually and could allow for the indefinite deferral of the realisation of animal rights. It also involves constraining the range of rights applicable to animals and limiting their application at present in a restrictive manner without a clear timetable for the expansion thereof.

What may be preferable then is to adopt a more expansive and structured approach where any violation of a right that is applicable to animals is recognised as being constitutionally suspect. In order to be constitutionally permissible, though, such a violation will have to pass the general limitation enquiry contained in section 36 of the Constitution which requires that certain criteria must be met for such a limitation to be justifiable. It is to this approach and its merits that I now turn.

3.2 General limitations

Section 36 (1) of the South African Constitution provides as follows:

The rights in the Bill of Rights may be limited only in terms of law of general application to the extent that the limitation is reasonable and justifiable in an open and democratic society based on human dignity, equality and freedom, taking into account all relevant factors, including

(a) the nature of the right;
(b) the importance of the purpose of the limitation;
(c) the nature and extent of the limitation;
(d) the relation between the limitation and its purpose; and
(e) less restrictive means to achieve the purpose.

This clause outlines the grounds upon which the limitation of a right may be justified. I now turn to certain features of this clause which may be of importance in the context of animal rights.

First, a law of general application is required for a right to be capable of being limited. I have already mentioned that this principle is rooted in the rule of law and partially gives expression to the grundnorm proscribing arbitrary treatment of individuals.86 Section 36(1) itself thus contains a commitment to these principles.

---

85 I have made a similar argument concerning the need to bring pragmatic factors into a determination of the core of the right to health care. See Bilchitz (n 39 above) 220 - 225.
86 See also Woolman (n 23 above) 34-54 - 34-67.
However, importantly, this requirement also means that any limitation on animal rights must be explicitly provided for in law. This entails that, in general, parliament would need to have engaged with the reasons for a particular practice that limits such rights and provide specifically in a statute for the regulation of that practice. This flows from the fact that rights may not be limited by laws which contain no express criteria concerning the circumstances where they can be limited and simply leave the limitation of rights up to the discretion of an official.87

This requirement that parliament engages with the limitation on animal rights could have several effects: On the one hand, it does involve parliament actively permitting certain violations of animal rights, which might appear to provide a stamp of approval for ethically unacceptable practices. Yet, on the other hand, South African legislation has to go through a public participation process which allows for members of the public to provide submissions in this regard. There would thus have to be transparent public engagement and openness as to the kinds of practices that occur in relation to animals, many of which often happen behind closed doors. Simply shedding light on some of these practices may be sufficient to demonstrate the lack of justification for the cruel treatment involved. In other cases, it may lead to the amelioration of the worst forms of suffering that flow from them. The process parliament follows would thus in itself require active deliberation concerning the manner in which animals should be treated. This in itself would involve a step forward and ensure that their interests are at least placed within the realm of public discussion.

Over time, and with education as to the interests and needs of animals, it is possible (perhaps even likely) that such public engagement could lead to wider protection for their rights. Indeed, recently in the United States, referenda were held in several states concerning the permissibility of practices such as sow stalls and veal crates being used in the meat industry.88 These referenda appeal directly to individuals in the political community and bypass the legislatures and lobby interest groups involved in agriculture. Almost without exception, these referenda have shown the public unprepared to tolerate the worst excesses of animal cruelty and laws banning these practices have as a result had to be passed. Whilst a similar direct form of democracy does not exist in South Africa, the requirement that parliament would have to pass laws that expressly limit animal rights could enable the public’s abhorrence at the worst forms of cruelty to enter into parliamentary deliberations through the public participative processes that have to take place surrounding legislation.

87 See Dawood (n 58 above).
Even a law of general application will, however, be evaluated in substantive terms to determine whether the limitation concerned passes constitutional muster. The factors outlined in the limitation’s clause are generally taken to require courts to undertake a broad proportionality enquiry. This means that the importance of particular rights for animals would have to be weighed up against the importance of any provision of law that seeks to limit those very rights. Any measure adopted must be connected to the important purpose it seeks to achieve and must be a particularly good means of realising the purpose. Of particular power in the limitation’s enquiry is the requirement that a measure must be considered not to be unduly restrictive of the rights under consideration in light of alternatives that could be adopted.

The application of this substantive enquiry would allow for some flexibility in relation to animal rights: Courts, for instance, could recognise that meat-eating is currently regarded as a legitimate purpose for limiting the rights of animals. On the other hand, not all abuses in the meat industry could be condoned simply on this basis: There would have to be an assessment concerning the relation between the limitation and the purpose as well as an evaluation as to whether less restrictive means could be used to achieve the purpose in question. In many cases, such a requirement could not be met.

Much of the suffering imposed on animals in the process of rearing them for food is unnecessary when considered in relation to the purpose. For instance, cattle are regularly castrated with no anaesthetics. Such mutilation, and the consequent violation of their basic rights to bodily integrity (and the right to be free from severe pain) could only be defended on commercial grounds since the existing practice speeds up the castration process without requiring analgesia or anaesthesia. It is unlikely that a court sympathetic to the importance of the rights of the animals concerned could accept such a commercial justification for this cruel practice: It would thus seem that less restrictive means could be employed to achieve the same result but prevent wholesale violations of animal rights. Similar points could apply to the sow stall and battery cages. Given the dire impact these practices have on animals, and the fact that they disproportionately harm animal interests given the purposes they seek to achieve and the alternatives that exist, the European Union has already

---

89 *Makwanyane* (n 8 above) para 149.
90 As has been mentioned, this could be recognised even if it is not objectively justifiable. Sows are kept in narrow metal-barred stalls that are barely bigger than the sow herself (known as a ‘sow stall’): She is prevented from being able to exercise and is unable to turn around for nearly four months at a time. This cruel practice is unnecessary and is being phased out by the European Union and in parts of the United States. See C Druce & P Lymberry ‘Outlawed in Europe’ in P Singer (ed) *In defence of animals* (2006) 124.
91 Laying hens are kept in battery cages that are so small they cannot stretch their wings, let alone walk or scratch the ground. These cages prevent most natural behaviours and lead to the degeneration of the birds’ bodies and deformities in their claws. About 4 700 million hens worldwide are kept in these conditions Druce & Lymberry (n 91 above) 128 - 129.
agreed to phase out many of these practices. It has, for instance, agreed to ban all cosmetic testing on animals from 2009, to ban the sow stall by 2013 and to prohibit the conventional battery cage by 2012 such that all production will use free-range methods.93

Such changes in farming practices would substantially improve the lives of a large number of animals without negatively impacting upon the lifestyles and well-being of human beings in South Africa. I have thus sought to provide some illustrations of the way in which the substantive features of the limitations clause could be used to allow for certain violations of animal rights whilst still ensuring that these rights make a difference to the lives of animals in South Africa.

A ground-breaking case decided in Israel illustrates the power of the principle of proportionality to find an adequate balance between human and animal interests.94 In that case, the Court was seized with a petition from an animal protection organisation known as 'Noah' requiring the Minister of Agriculture to issue regulations prohibiting the force-feeding of geese for the production of foie gras. The key statute involved in this case was the 1994 Cruelty to Animals Law that sought to prohibit cruelty towards vertebrate animals. As background to the decision, it is important to understand that the foie gras industry in Israel had existed for 40 years and was the fourth largest in the world.95 Foie gras is the fatty liver of a goose or duck and can only be produced if birds are force-fed. Such animals are force-fed three times a day with an air pump through a long metal tube: The tube itself can cause bruising, lesions and perforations of the oesophagus. After a month of such force-feeding, the birds' livers swell up to about 12 times their natural size which makes walking and breathing difficult. The pre-slaughter mortality rate for foie gras production is up to 20 times the average rate on other bird farms.96

In a remarkable decision, the majority of the judges ruled in favour of banning the force-feeding of geese. Justice Tovah Strasberg-Cohen recognised that the state of Israeli law required a balancing of animal and human interests to take place.97 Her inquiry involved three steps to decide whether 'unnecessary suffering' was caused to animals in the foie gras industry: The first step considered an evaluation of the purpose of the law

93 Druce & Lymberry (n 91 above) 126 130. On the ban on testing cosmetics on animals, see http://www.cosmeticsdesign-europe.com/Products-Markets/EU-refuses-bid-to-lift-ban-on-animal-cosmetics-testing.
95 Sullivan & Wolfson (n 88 above) 143.
96 Sullivan & Wolfson (n 88 above) 143.
97 It could be argued that such a balancing process is required even in South African law as it stands by the notion of ‘unnecessary suffering’ that is contained in the Animal Protection Act 71 of 1962 in, eg, sec 2(1)(r).
and whether it represented a worthy social purpose; the second concerned an evaluation of the relationship between the means adopted and the purpose in question; and, finally, a determination had to be made concerning proportionality, namely whether there was a correct balance between the suffering caused to the animal and the purpose and means adopted that caused such suffering. These tests, the justice recognised, were similar to those used in the context of limiting human rights.

In considering the purpose of the regulations, Justice Strasberg Cohen saw their aim as being to minimise the suffering of the geese. She found that it is quite clear that the practice of force-feeding causes the geese great suffering and that existing regulations had not minimised that suffering. Secondly, in considering the means adopted, she recognised that they were seeking to allow for the maintenance of the practice of foie gras production. Whilst the production of food was more important than pure entertainment, Justice Strasberg Cohen stressed the importance of distinguishing between ‘luxury’ and ‘necessary’ food products. The ‘luxury’ nature of foie gras was a factor that had to be considered in the balancing process when considering the grave harm that was caused to the geese. Similarly, the nature and weight of existing agricultural interests in maintaining foie gras production had to be considered but was not decisive.

Finally, the benefit obtained from the production of foie gras had to be weighed against the harm caused to the geese: Justice Strasberg-Cohen found that the agricultural interests had been given too much weight and that of the geese too little. She found when balancing the benefits and harms, that the current regulation by allowing the continuation of force-feeding of the geese failed the test of proportionality.

Justice Eliezer Rivlin concurred with Strasberg-Cohen. He was sympathetic to the important claim that animals cannot be made to suffer simply for gastronomic pleasure or profit, and again, considered the matter as one of proportionality between benefits and harms. He recognised that animals 'possess a soul that experiences the feelings of happiness and grief, joy and sorrow, affection and fear ... All would agree ... that these creatures feel pain inflicted upon them by physical injury or by violent intrusion into their bodies.' The problem with justifying the continuation of the industry simply for the livelihood of those who raise
geese and the gastronomical pleasure of those who eat foie gras is that it has a price – ‘and the price is the degradation of man’s own dignity’.  

Justice Asher Grunis wrote the minority opinion and found, similarly, that a balancing process had to take place. He placed greater stress on the agricultural interests concerned and, in particular, the livelihood of farmers in the industry.

In this light, he found that the suffering of the geese was proportionate to the benefit of producing foie gras. Such a judgment could change over time and he was of the view that ultimately ‘the current arrangement should not continue indefinitely, for the suffering of the geese should not be ignored’.

The order made by the majority ultimately invalidated the regulations that allowed foie gras to be produced although it suspended the order for 18 months. The suspension gave the Ministry the opportunity to put alternative regulations in place that would ban the industry but make provision for those who had made their livelihoods through the production of foie gras. The order had to be suspended several times given executive inaction until it eventually took effect in July 2005. The practice of force-feeding geese, however, continued until February 2006, when the High Court made another order forcing the state to enforce the judgment within two months.

This case is a very important precedent in that it demonstrates the importance of a balancing approach when faced with conflicting interests between animals and humans. First, all the justices clearly recognised that animals have interests which require protection and may be subject to severe suffering. Whilst not deciding the question on the ground of ‘rights’ (given the particular challenge), the reasoning is analogous to that which would occur in terms of our limitations clause.

Secondly, a substantive enquiry had to be conducted into the value of foie gras production to society and the harms caused to the geese. Ultimately, the severe suffering caused from this practice could not be justified for the selfish gastronomic benefits obtained by some, and even to ensure individuals were able to maintain their employment. This suggests the possibility that, in certain instances, human interests may be outweighed by animal interests. In the South African context, it is necessary to recognise animal interests at the constitutional level in order to conduct an evaluation in this manner and to avoid human interests – such as the freedom of trade and occupation – in all cases outweighing animal interests.

105 See Noah (n 94 above), the Opinion of Justice Asher Grunis para 31.
106 As above.
107 As above.
108 As above.
Finally, the Court in this case sought to give due consideration to the position taken by the elected branches of government. It, however, recognised the importance of the role that the judiciary had to play in evaluating whether the government had in fact achieved a position that proportionately balanced the competing interests concerned. In South Africa, this would require the judiciary to be prepared to recognise the applicability of constitutional rights to animals and to engage in substantive judicial review in relation to the measures adopted by the legislature and executive in relation to them. Importantly, as this analysis has shown, constitutional law has the tools to enable courts to achieve a sensitive balancing of interests through applying the tried and tested principles contained in the limitations clause.

4 Conclusion

This paper has sought to develop the implications of transformative constitutionalism for non-human animals. It began by outlining the interpretive method that should be adopted for a Constitution such as that in South Africa: The approach needs to be purposive and must consider the historical context. In particular, that context must make clear the values that are to be realised through the constitutional text and interpretations that are to be avoided.

The backdrop to the Constitution showed that the new order represented a reaction against the systematic and arbitrary denial of equal treatment to black people simply on the basis of their race. This led to the recognition of a fundamental grundnorm underlying the new constitutional order which prescribes that there be no arbitrary discrimination against any individual without a reasonable justification. The argument was provided that discrimination on grounds of species is just as arbitrary as any other form of discrimination. As a result, the grundnorm of the South African Constitution should be seen as extending to species. This would mean interpreting the Constitution so as to prohibit discrimination on grounds of species and to recognise the entitlement of non-human animals to equal treatment. It was necessary, however, to find an interpretation of the Constitution that accords with this ethos.

Part II involved developing an interpretation of the Constitution in light of the fundamental grundnorm and its implications. The discussion sought to demonstrate that constitutional rights could be applied quite naturally to non-human animals give that they could be included within the designation 'everyone' in the Bill of Rights. It was argued that the value of human dignity did not provide an insurmountable barrier to the recognition of animal rights and that animals must be recognised as having some worth at the constitutional level. The value of equality and the jurisprudence of the Constitutional Court was then analysed and I sought
to show how they could provide support for the extension of rights to animals. I also provided a number of arguments as to why we should adopt a presumption in favour of recognising that non-human animals have equal worth and why we should be suspicious of judgments of differential worth. Equal worth does not, however, require identical treatment and I outlined the way in which equality could be applied beyond the human species.

Finally, I considered the difficult question where human and animal rights conflict. In particular, the problem arises in that human society, whilst extremely powerful with respect to animals, is not yet ready to recognise the full implications of animal rights. I outlined two strategies to address this problem both rooted in concepts based in the South African Constitution. The first involves the notion of ‘progressive realisation’ which I argue can be adapted to the context of animal rights. It would involve a recognition that the full realisation of such rights may not be possible now though a minimum core of such rights must be protected. Progressive steps must be taken towards the full realisation of such rights over time. Secondly, I argue that the general limitations clause of the Constitution provides important principles that should be used when considering limits on animal rights and dealing with circumstances of conflict. In particular, the proportionality test provides a rubric for considering the differing interests and their respective weight. The ‘less restrictive means’ requirement would also be of great importance to ensuring that the extension of rights to animals makes a difference to animal lives without going too far beyond what society is prepared to accept in this regard.

The interpretation of the Constitution I have defended is one that draws out the deepest principles of the new constitutional order and applies them in a way that would perhaps surprise many in our society. Yet, this should not be a reason to refuse to acknowledge the power of the argument or the interpretation provided. The Constitution was not designed to entrench the status quo: It was designed to transform our society into something different, to allow ourselves to imagine a social order based on firm principles of justice and equality. Leaving out animals would be completely inimical to this ethos and perpetrate unjustifiable discrimination in the new order. It is time, both philosophically and in concrete terms, to take the logic of fundamental rights to its ultimate conclusion: to recognise that all beings with interests – irrespective of the classifications we give them, or the physical make-up of their bodies – are entitled to the protections offered by the rights in the Bill of Rights.
CHAPTER
ANIMAL RIGHTS AND THE INTERPRETATION OF THE SOUTH AFRICAN CONSTITUTION

Thaddeus Metz*

1 Introduction

In a recent work, Bilchitz advances two important claims about South African constitutional law.¹ One claim is that, because animals have a dignity that demands respect, the Constitution ought to be amended so that animals are explicitly deemed to be ‘persons’ and hence entitled to protection under the Bill of Rights. Of course, not all the rights in the Constitution’s second chapter would apply to animals; clearly, if animals have dignity, respect for it would not require according them a right to vote. However, according to Bilchitz, it would be apt to consider animals expressly to have rights of bodily integrity and freedom of movement, among others.

Bilchitz’s second claim is that, in the absence of a constitutional amendment that would ground animal rights in a plain reading of the text, the Constitution is best interpreted as already including them. In his view, a proper reading of the text would extend legal personhood (whether natural or juristic) to animals since they have an inherent dignity. However, Bilchitz maintains that the constitutional rights of animals should be systematically limited by the doctrine of progressive realisation. That is, since it is currently impossible to enforce animal rights fully, which would in principle prohibit eating animals merely for the taste, a minimum core of protection from more glaring and easily avoidable forms of cruelty, such as castration without anesthetic, should be enforced, with additional rights being enforced over time as it becomes more feasible.

I find the broad outlines of the arguments put forward by Bilchitz quite convincing. Many moral philosophers and professional ethicists believe that the establishment of certain duties toward animals is one of the easiest practical disputes to resolve. Virtually no one believes it is permissible to set a cat on fire merely for the thrill, and nearly everyone believes that the best explanation of why this is wrong has something to do with the effect on the cat. At least some animals have a worth in their own right that merits moral treatment, and it is not an enormous leap from this claim to the idea that certain forms of wrongdoing with regard to animals should be legally prohibited. Bilchitz’s suggestions about precisely how constitutional principles and rights might be understood to apply to animals are revealing, fascinating and worth taking seriously.

I do, however, question some of the specifics of Bilchitz’s view about the way the law ought to accord rights to animals. In this article, I focus on Bilchitz’s claim that, in the absence of an amendment expressly recognising animal rights, the South African Constitution is best read as already including them, albeit limited by the principle of progressive realisation. Whereas Bilchitz maintains that constitutional justices should not fully enforce animal rights at present because it is impossible, I provide two reasons to believe that they should not fully enforce, and perhaps not even recognise, animal rights at present because it would be all things considered unjust. I am not sure whether these two arguments are sound. I am, however, certain that they need to be addressed before having conclusive reason to favour a reading of the Constitution, as it stands, as embodying animal rights.

2 The argumentative strategy

Bilchitz claims that the Bill of Rights is best read as applying to many animals, although many of these rights are properly limited by the doctrine of progressive realisation. In his view, constitutional justices would be correct to deem animals to be legal persons that have a minimum core of rights that may not be infringed, for instance, rights not to be subjected to cruelty when it would impose trivial costs on human beings, such as the right of an animal not to have its genitals removed without a painkiller. However, many other animal rights are incapable of being enforced, given current sensibilities in South Africa, for example, the right of an animal not to be eaten merely because it tastes delicious. These currently unenforceable rights, Bilchitz maintains, should instead be realised over time as the culture changes and, more specifically, as the government promotes changes in culture so as to enable more animal rights to be enforced.

I agree with Bilchitz that animals have a moral worth that merits respect not merely at the private, individual level, but also at the public, legal one, at least under certain conditions. However, it is not clear to me that the moral status of animals is such as to merit interpreting the Constitution as it stands to
grant legal personhood to animals and, hence, protection by the Bill of Rights. My reasons for suspecting that it is impermissible to interpret the Constitution in this way do not rest, as one might think, on a ‘conservative’, ‘originalist’ or otherwise ‘passive’ theory of constitutional interpretation. In fact, I share the broadly purposive or natural law approach to reading legal texts that Bilchitz invokes. However, I find attractive a particular version of naturalism that apparently entails that it would be wrong for constitutional justices to apply the Bill of Rights to animals in the absence of a constitutional amendment; or, at least, my value-laden interpretive philosophy, which accords with the moral judgments of many jurists and philosophers of law, appears to entail that the reason not to fully enforce the Bill of Rights with regard to animals in the absence of an amendment is not (merely) that it would be impossible, but that it would be a greater defect of political morality to fully enforce such rights than not to enforce them. Again, my strategy is to grant Bilchitz that, given the moral worth of animals, there would be some injustice in failing to read the South African Constitution in a way that accords them protection from cruelty and other mistreatment, but to consider whether a more weighty injustice would be done if it were so read.

Note a few limitations of the argumentative strategy that I will explore. First, I grant throughout that many animals have a moral status, and, indeed, even a dignity of a sort. The arguments I discuss are strongest, though, if animals generally have a moral status that is less than what persons have, a position that I shall argue for in what follows, but that I lack the space to defend with philosophical conclusiveness. Second, while the principles of justice that I appeal to are meant to be universally applicable, they have particular implications in light of South Africa’s history, and could have different implications in other social contexts. Third, I set aside the issue of whether a constitutional amendment to accord rights to animals would be permissible or not. The reasons I discuss for not interpreting the Constitution as already including animal rights probably provide some ground not to amend the Constitution to include them; however, I am unsure of whether it is definitive or not. I suspect not, at least in the case where parliamentary ratification of an amendment is consequent to the development of a substantial, favourable public view on the matter. However, I say no more about the bearing the arguments I make for how to interpret South African law might have on how to make law in this country.

3 Formal justice

Considerations of what is sometimes called ‘formal justice’ provide reason to think that the best interpretation of the Constitution as it stands entails that it would be wrong to fully enforce animal rights, and perhaps even to recognise them at all. Formal justice is, roughly, a matter of an agent consistently applying principles that it believes to be just, where substantive justice, in contrast, is a function of the content of the principles applied. In a
criminal trial, it would be substantively unjust to impose a severe penalty on someone guilty of a trivial crime, and it would be an additional, formal injustice to impose such a principle selectively, say, only on those who committed crimes on a Tuesday.

Although Dworkin does not use the language of ‘formal justice’ (that I recall), his naturalist theory of interpretation essentially appeals to it, and I will use his theory to ground my objection to Bilchitz. According to Dworkin, constitutional justices ought to read the text in light of the most justified principles of political morality that make sense of the recent history of their legal system as a whole. They are to appeal to defensible principles of substantive justice, but not necessarily the most defensible considered on their own from a philosophical point of view. Instead, judges are to find the most defensible principles of substantive justice that ‘fit’ the legal system in which they operate, that is, that adequately entail and explain a wide array of judgments, norms and practices of contemporary law as a whole. For example, Dworkin has us imagine that a judge ascertains that the philosophically most defensible distribution of economic wealth is socialist, but that she lives in a legal system that is thoroughly capitalist. Dworkin maintains that such a judge would have some reason of substantive justice to render socialist verdicts, but more reason not to, as the principles of justice to which she appeals must not be overly discontinuous with her legal context.

Dworkin provides several reasons for the ‘fit’ criterion of legal interpretation, with the most interesting and powerful ones able to be placed under the heading of ‘formal justice’. The basic idea is that a judge in a thoroughly capitalist system who ruled socialistically would be failing to uphold her duty to assist the legal system in consistently applying principles that it maintains are just. If one does not see any immorality in ‘formal injustice’, consider two arguments Dworkin advances for finding it so, one self-regarding and one other-regarding.

In terms of other-regarding considerations, Dworkin maintains that a judge would fail to treat as equals those subjected to her idiosyncratic, but perhaps substantively just, decision. When there are great ruptures in judicial interpretation, the government fails to speak with one voice and thereby unfairly treats one group of citizens according to one standard, and another group according to another one.

With respect to self-regarding matters, Dworkin believes that the consistent application by government of principles that it deems just is necessary in order for it to exhibit the political virtue of integrity. If judges appealed to whatever principles of justice they found most substantively

---

2 R Dworkin ‘‘Natural’’ law revisited’ (1982) 34 Univ Florida LR 165; R Dworkin Law’s empire (1986); R Dworkin Freedom’s law (1996).
3 Dworkin Freedom’s law (n 2 above) 11.
4 See especially Dworkin Law’s empire (n 2 above).
justified, without consideration as to whether the principles cohere with recent legal practice, then the state (as a moral agent distinct from the individuals who compose it) would fail to be principled, would act haphazardly, and would probably even count as hypocritical. If we want political institutions to manifest the virtues of honesty, gratitude and remorse, then it seems apt to want them to exhibit integrity as well.

The implications of Dworkin’s attractive theory of interpretation for Bilchitz’s argument should be clear. Even if Bilchitz were correct that ideal principles of substantive justice require applying the Bill of Rights to animals, it would not necessarily follow that judges, all things considered, have reason to interpret the Constitution in that manner. They would need to factor in considerations of formal justice, and there is of course strong reason in the case of animal rights to think that recognising them at the constitutional level would be seriously discordant with South Africa’s recent legal history. After all, if legal personhood at the constitutional level were not such a radical break, Bilchitz’s claims would be of less jurisprudential and academic interest.

Of course, Bilchitz does not recommend the full enforcement of animal rights, claiming only that a minimal core of protection should be enforced, with other protections being increasingly adopted over time as they become feasible. But I suggest that Bilchitz has provided the wrong reason for limiting animal rights, or at least not all the relevant reasons. Bilchitz rejects the full enforcement of animal rights because it cannot be accomplished. The doctrine of progressive realisation, as Bilchitz says of it, ‘recognises that the full realisation of these rights may not be possible at a particular point in time’.\(^5\) Of course, if an action is not possible, then it follows that an agent lacks any reason either to do it or not to do it. Whereas Bilchitz is saying, in effect, that there is not reason to fully enforce animal rights, the view I am considering is that there is reason not to fully enforce them: Even if it were possible to fully enforce animal rights, a justice might be wrong to read the Constitution in a way requiring that, as doing so would violate principles of formal justice.

The natural reply for Bilchitz to make at this point is to contend that his interpretation of the Constitution would not infringe Dworkin’s invocation of formal justice, as it would bring out deep principles of justice that, with respect to South Africa, ‘show the history of judicial practice in a better light’.\(^6\) Specifically, Bilchitz maintains that just as the Constitution clearly forbids racism because of its arbitrariness, so the Constitution should be read as forbidding speciesism for the same reason. However, the burden of Dworkin’s hermeneutical approach is that those reading the Constitution need a construal of the requirement to avoid arbitrariness

---

\(^5\) Bilchitz ‘Moving beyond arbitrariness’ (n 1 above) 70.

\(^6\) Dworkin “‘Natural’ law revisited’ (n 2 above) 169.
that adequately coheres with recent judicial history, something Bilchitz’s construal probably does not. For a judge in today’s South Africa to apply the Bill of Rights to animals and to treat speciesism as on a par with racism would be akin to a judge rendering socialist verdicts in a capitalist society.

4 Compensatory justice

 Whereas the previous argument maintained that deciding whether to constitutionally recognise the legal personhood of animals involves a trade-off between substantive justice and formal justice, the present one maintains that distributive justice for animals in South Africa might come at the cost of compensatory justice for people and, in particular, for Africans. Bilchitz appeals to principles of distributive justice, contending that an ideal distribution of liberties, resources, restrictions and burdens entails that the Constitution is properly interpreted as according rights such as bodily integrity and freedom of movement to at least some animals. The other sort of justice that I invoke is compensatory, a subset of non-ideal principles indicating how to respond to past violations of ideal principles of justice. Principles of compensatory justice tell us how rightly to deal with wrongful behaviour, and, specifically, to do so by effecting restitution in some way. A requirement to make up for wrongful damage done is widely recognised as being a suitable aim for political institutions, particularly when they themselves have done the wrongful damage – hence, to give just one example, the TRC’s call for state reparations to victims of apartheid-era political crimes.

If the Constitutional Court were to adopt Bilchitz’s recommendations about how to read the Constitution as it stands, it would not merely fail to help effect restitution among the previously most wronged people in South Africa, but would likely retard achievement of that aim. One of the major injustices of apartheid took the form of the forcible eclipse and denigration of African cultures. There is debate among political philosophers about the precise respect in which this was an injustice – for instance, some would say that culture itself is a good of which Africans were robbed, while others would contend that the problem is the self-esteem that was foreseeably reduced via the destruction of culture. One need not settle that debate in order to recognise that a plausible way for the state to repay those whose cultures it destroyed would be for it to foster their cultures. Concretely, this could take the form of: supporting the study of traditional African societies at public universities, funding local museums that would protect and showcase physical artifacts, paying people to discover and interpret intangible heritage such as ideas associated with talk of ‘ubuntu’, digitising the narratives of oral peoples, employing African

---

7 For a classic source of the distinction between ideal and non-ideal principles of justice, see J Rawls A theory of justice (1971).
languages in publicly sponsored discourse, and so on. Akin to these policies would be the practice of giving some (not necessarily conclusive) weight to characteristic African values when making legal decisions.

Now, it is typical of (Southern) African culture not to accord animals a dignity, that is, a superlative intrinsic value, or at least not one that would approximate that of human beings and warrant legal enforcement. In the Southern African (and more generally sub-Saharan) region, the maxim taken to summarise morality is usually translated as ‘A person is a person through other persons’. This is Motho ke motho ka batho babang in Sotho-Tswana, and Umuntu ngumuntu ngabantu in the Nguni languages of the Zulu, Xhosa and Ndebele. The basic idea of the maxim is that one becomes a moral person or lives a genuinely human way of life, manifesting ‘botho’ or ubuntu, just to the extent that one lives in community with other people.8 One need not delve any deeper into the essentials of this ethical worldview to see that the relevant beings with which to relate communally are solely persons, a group that is usually held to include ancestors (and, in some traditional societies, spirits who are not yet born), but does not include animals.

Of course, it does not follow that, for a southern African morality, one may treat animals or the rest of nature any way that one pleases. Instead, person-centred reasons are usually given for thinking that it would be wrong to be cruel or otherwise treat animals in intuitively immoral ways.9 For example, one routinely finds the rationale that, since everything in the world is interdependent, treating persons well requires not exploiting the natural world. For another example, to live communally with ancestors can require protecting land that they are deemed ultimately to own, or respecting animals that are considered totems.

However, none of these recurrent rationales for not interfering with animals appeals to the dignity or even moral worth of the animal.10 In a large majority of Southern African cultures animals are routinely eaten for the taste, slaughtered to pay tribute to ancestors, and worn for ornamentation. To interpret the Constitution in a way that forbids these practices, even if subject to progressive realisation, is therefore not merely to fail to uphold African cultural practices, but also to judge them negatively.

---

9 Many of the following rationales can be found in the sixth part of MF Murove (ed) African ethics: An anthology of comparative and applied ethics (2009).
10 As one would expect given the diversity of views in the sub-Saharan region, one can find some grounds for according moral status to animals. There is a strain of African thinking that can be read as maintaining that community is to be fostered not merely with persons, but also with some other beings in nature, on which see K Behrens ‘Exploring African holism with respect to the environment’ (2010) 19 Environmental Values 465. Even here, though, few Africans would deem such a perspective to forbid eating or wearing animals, let alone to warrant the legal prohibition of these practices.
and to suppress them even more. Indeed, this is the way that the African National Congress (ANC) judged recent attempts by Animal Rights Africa (ARA) to use the law to prohibit *ukweshwama*, a Zulu practice in which adolescent men bare-handedly kill a bull in order to signal the time to harvest, to honour the strength of their king, and to express gratitude to ancestors.

For hundreds of years, the apartheid and colonial systems sought to relegate the cultural and traditional practices of the majority of South Africans to the humiliating levels of lowliness. Millions of black people, Africans in particular, were dictated to on how to conduct their cultures and a foreign way of living was imposed on them ... (W)hile there is no question about the paramount role the courts play in our constitutional democracy, they might not be the appropriate platform for resolving complex and sensitive matters such as *ukweshwama*. The ARA's decision in this regard sets a worrying precedent in that those with huge financial resources can dictate to others how their age-old traditional customs should be conducted, as was the case during the apartheid and colonial eras.11

Since there is a weighty duty on the state to make up for the losses of culture it has been responsible for, justices might have, all things considered, reason not to add to more legal prohibition of the culture with which many Africans identify.

As it turns out, this reasoning is not far from some of the principles invoked in the High Court's decision with regard to the bid on the part of animal rights activists to stop *ukweshwama*.12 Although this Court did not refuse outright to hear the petition, as the quotation above suggests the ANC would have preferred, it did find in favour of representatives of the Zulu people, and did so in part on the grounds of historical discrimination.13 The High Court says:

> From a historical perspective applications of the present are nothing new and are symptomatic of an intolerance of religious and cultural diversity ... The traditional African form of culture, religion and religious practices ... were historically often discriminated against and in some instances its followers were persecuted and punished ... [The applicants have] called into question the legitimacy of the religious and cultural practice and offended the members of the Zulu nation who are now called upon to justify their beliefs and

---


13 Additional reasons from the High Court include the absence of evidence that the killing of the bull is consequent to great suffering, and a concern about violence erupting if the killing were legally prohibited.
The reference to past injustice suggests that the Court’s reasoning is not merely that the Zulu nation’s right to culture outweighs the ARA’s right to conscience, an argument based on distributive justice that is worth considering, but that I do not focus on here. What I instead highlight here is that one apparent aspect of the Court’s reasoning is the idea that prior discrimination against the Zulu people’s culture provides some reason to protect it now, or at least for the law not to interfere with it.

In response, Bilchitz would no doubt suggest that not all elements of a culture are worth retaining, which is, of course, true. It is right, for example, to forbid *muti* killing on constitutional grounds, despite being a part of some indigenous cultures. If culture is not sufficient to justify the killing of an innocent child for the purpose of obtaining dubious medicinal elements, then, so the response would go, culture is not sufficient to justify the killing of an innocent animal for the purpose of taste, decoration or religion, even if there are reasons of compensatory justice to foster the culture of those who have been robbed of it.

Implicit in this response, most likely, is the assumption that the killing of an innocent person and the killing of an innocent animal are morally on a par. Indeed, Bilchitz explicitly maintains that the dignity of an animal has an ‘equal moral weight’ to that of a person. While I accept that animals matter morally in their own right, so that it is a grave wrong to inflict unnecessary cruelty on them, I find it implausible to think that they have as much moral status as characteristic persons. A variety of uncontroversial judgments are evidence for this view. For example, if the reader and I were starving along with a pig, and if I had a gun, it would be permissible for me to shoot the pig to feed us, but impermissible for me to shoot the reader to feed myself and the pig. The best explanation of such a judgment, I submit, is that pigs have a lower moral status than persons. Similar cases abound. Imagine you are driving a bus and must choose between running over a person and a cat; *ceteris paribus*, you must strike the cat, the most plausible explanation of which is that the cat is not as morally important as the person. Or consider a case in which you must choose between killing a pit bull terrier or letting it bite a child. Such forced trade-offs are rare, to be sure. But thinking about what it would be right to do in such cases, and why, reveals much about our basic views of moral standing between animals and persons as unequal.

Differential moral status also accounts best for uncontroversial judgments about how to treat beings that have already been killed. If an

---

14 Smit v Zwelithini Kabhekuzulu (n 12 above) 13.

15 Bilchitz ‘Does transformative constitutionalism require the recognition of animal rights?’ in Woolman & Bilchitz (n 1 above) 196.

16 It does not follow that it would be permissible to eat the pig in the situation where sufficiently-nourishing, pleasant-tasting, vegetarian food were readily available.
animal has been killed for whatever reason, many find it permissible not to let it go to waste; even many vegetarians would find something respectful in the stereotypical native American practice of using every part of a buffalo, once it has been brought down. In stark contrast, such a practice applied to human persons would be horrific. Consider a Nazi thinking, 'Well, we have already killed this Jew, and so may as well make the best of it by using his hair to stuff pillows, fat to make soap and bones to fashion buttons'. If animals and persons had the same moral status, our reactions to the native American and Nazi cases of posthumous treatment would be the same, but they are not.

I have not provided a theoretical specification of the property in virtue of which persons have a higher moral status than animals, something I lack the space to do here, and something I need not do in order to mount a serious challenge to Bilchitz.

I have provided strong reason to think that persons and animals have a differential moral standing, even if I have not said why, and if indeed animals have a lower moral status than persons, then it is likely that the state should prioritise the latter’s interest in compensatory justice with regard to cultural wrongs.

Bilchitz can reasonably reply, at this point, that even supposing there is unequal moral status between persons and animals, the moderate interest of a human being in culture is outweighed by the urgent interest of an animal in its life. Instead of seeking to rebut this point, I note that Bilchitz’s approach to constitutional interpretation would, in fact, unreasonably require trading off the urgent human interest in human life for the sake of animal lives. Progressive realisation of animal rights would mean that the state would have to spend resources to move society closer to the end in which they are fully recognised. Furthermore, constitutional recognition of so-called ‘negative’ rights against cruelty and the like often requires substantial amounts of time and money, not merely from the state, but also from the rest of South African society to whom such rights would horizontally apply. Now, these are scarce resources that, in principle, could go toward paying Africans back in other, more material ways for past injustices. When literally many hundreds of thousands of Africans die each year in this country from diseases and injuries that a dysfunctional healthcare system cannot treat, and when a poorly developed educational system leaves millions of Africans to meaningless, undignified and unhappy lives of unemployment and severe poverty, it would express disrespect for them if the state and others in society were

---

Animal rights and the interpretation of the South African Constitution

219
to expend severely limited resources on the urgent interests of animals. Or, at the very least, it might reasonably be perceived to express such disrespect, akin to sacrificing an African’s life for the sake of saving a pig.

5 Conclusion

I have argued that, even supposing that Bilchitz is correct that substantive principles of distributive justice entail that animals warrant constitutional protection, there are other, potentially weightier forms of injustice that would probably be done by interpreting the Bill of Rights as already applying to animals, namely, formal injustice and compensatory injustice. Formal injustice would apparently result from such a reading of the Constitution in that the South African state would fail to speak with one voice upon according legal rights to animals. Compensatory injustice would likely result from such a reading in that the law would not only suppress facets of culture that many Africans deem important to their self-conception, but also require spending scarce resources on animals that could have gone toward saving African lives and livelihoods. Human rights not to be denied life-saving healthcare and not to suffer from poverty continue to be violated as a result of apartheid era policies. If the state must choose between acting for the sake of the urgent interests of animals and those of humans, humans must take priority, even assuming that animals have a kind of dignity that morally forbids harming them in our private lives.

I am not certain that interpreting the Constitution as already according rights to animals would violate principles of formal and compensatory justice. Perhaps doing so would justifiably infringe them, or, less plausibly, maybe doing so would be consistent with them. My aim has been to indicate some moral and legal issues that need to be thought through before making a conclusive judgment about whether to read the Constitution in the ways Bilchitz proposes. Even if he is correct that principles of distributive justice entail interpreting the Constitution’s Bill of Rights as it stands as applying to animals, this consideration must be weighed up against principles of formal and compensatory justice that appear to conflict with such an interpretation.\(^\text{18}\)

\(^{18}\) For helpful comments on an earlier draft of this article, I thank David Bilchitz, an anonymous referee for \textit{SA Public Law}, and participants in a seminar sponsored by SAIFAC.
IS THERE A DIFFERENCE THAT MAKES A DIFFERENCE BETWEEN UBUNTU AND DIGNITY?

Drucilla Cornell*

1 Introduction

In South Africa, both legal and philosophical scholarship frequently draws important contrasts between dignity and ubuntu. In this essay I want to challenge some of the assumptions that underlie these sharp contrasts.

I will also argue, however, that there is a difference that makes a difference between ubuntu and dignity, and that this difference is important in the continuing struggle for a truly ‘NEW South Africa’, particularly in efforts to challenge the neoliberal policies of the African National Congress undertaken on a daily basis by on-the-ground movements. Legally, ubuntu is important, as would be expected, in the arena of socio-economic rights – but not just there. Both ubuntu and dignity have been deployed in the battle to maintain an ‘outside’ to the capitalisation of all human relationships. I will argue that ubuntu, rather than dignity, may well serve as the more adequate opposition to the drive to turn all human relationships into commodities and cash them out for their value in the marketplace.

Think for example of the Dikoko decision1 discussed at some length in ‘Recognition of ubuntu’.2 There, Justices Sachs and Mokgoro both rely on ubuntu to argue that what was at stake in the violation of reputation was not money but rather the reconciliation of the individuals involved – and indeed, that the long accepted idea that money is the appropriate award for personal injury actually exacerbated the damage to the relationships between the parties to the lawsuit.

---

* Professor of Political Science, Women’s and Gender Studies, and Comparative Literature, Rutgers University, New York, USA. Extraordinary Professor, University of Pretoria. Visiting Professor, Birkbeck College, University of London.

1 Dikoko v Mokhatla 2006 6 SA 235 (CC).

But I am getting ahead of myself. First, let us contrast dignity and *ubuntu* with a broad brush, as they have often been painted in the literature. Dignity is usually associated with individual human rights, with a strong emphasis on autonomy and personhood, while *ubuntu* is associated with communalism and such virtues as loyalty and generosity. When it comes to constitutional law in South Africa, however, what they have in common is that neither seems easily reconcilable with reigning notions of legal positivism. That said, *ubuntu* has been more persistently attacked as supposedly untranslatable into a judiciable principle of constitutional law. This is in part because dignity is clearly more familiar to Westerners and more easily accepted by those scholars in South Africa trained in Germany or in German law. For modern legal systems such as those of Germany and Israel, dignity is a foundational principle. Article One of Germany’s 1948 post-war Basic Law makes human dignity an inviolable right and one that is to inform the spirit of the legal system as a whole. The 1992 Israeli Basic Law makes human dignity a fundamental moral ideal and a right within the Israeli legal system. In international law, meanwhile, the 1948 Universal Declaration of Human Rights establishes the dignity of all persons as the basis for freedom and justice in the world.

Dignity then, at least since World War II, has a rich legal history – one that has certainly played a role in the development of the dignity jurisprudence in South Africa, and many of us have argued that the depth and breadth of the dignity jurisprudence is one of the most significant contributions of South African constitutionalism. Can the same be said of *ubuntu*? The answer is yes and no. Certainly one does not find *ubuntu* in the Universal Declaration of Human Rights. On the other hand, if one puts *ubuntu* in the rich tradition of African humanism and socialism – and I think it should be connected to this intellectual heritage – then *ubuntu* (and similar African ethical principles such as *Ujamaa* in Tanzania) has a very rich history indeed, and one that places such principles at the heart of the creation of a post-colonial Africa.

Given the troubled history of constitutionalism in Africa, the legalisation of such principles has been mixed, yet they clearly play a major role in defining the *grundnorm* of an ethic for a new society. A *grundnorm* is Hans Kelsen’s word for the grounding moral or ethical principle that undergirds not only the legal system but the society as a whole. Such a *grundnorm* was explicitly defended by many of the African leaders after the revolutionary struggles for independence through principles of African Humanism such as *Ujamaa* and *ubuntu*.

It is an honour to present this essay in the company of Justice Mokgoro who has played such an important role in what Sam Fuller and I have named the ‘recognition’ of *ubuntu* by the Constitutional Court. Justice

---

3 H Kelsen *General theory of law and the state* (1945).
Mokgoro’s important judgments are discussed at length in ‘The re-recognition of ubuntu’. Famously, ubuntu remains in the final Constitution of South Africa as merely the shadow that justifies a Truth and Reconciliation Commission. However, in a series of interviews and lectures, Justice Mokgoro has argued that, even if this is the case, ubuntu can and should be interpreted as the ethical law of the entire Constitution.

If one were to interpret ubuntu as the ethical law of the new South Africa, then it would be ubuntu that would ground the constitutional grundnorm of dignity and not dignity that calls for the re-cognition of African humanist principles such as ubuntu. Note that I have used the word African because ubuntu or ubuntu/botho, even as they are words in South African languages, should be understood as part of the rich intellectual heritage of African humanism.

2 Kantian dignity

But let us first turn to the most sophisticated European defender of dignity, Immanuel Kant, so that we can contrast the Kantian justification of dignity with a rethinking of ubuntu, remembering also that in South Africa, Kant has been explicitly incorporated into constitutional jurisprudence. Former Constitutional Court justice, Justice Ackermann, has powerfully defended Kantianism as one important secular justification for the understanding of dignity as an ideal attribution of persons such that all persons have intrinsic worth. His constitutional judgments boldly suggest that the new South Africa should be understood to aspire to the great Kantian ideal of the Kingdom of Ends.

For Kant, a human being is of incalculable worth and has dignity precisely because through our practical reason we can potentially exercise our autonomy and lay down a law unto ourselves, which for Kant is the moral law or the categorical imperative. Kant is not arguing that we actually do exercise our autonomy most of the time in our daily lives. Nor is he defending autonomy as some kind of truth about how we actually are – a common misreading of his argument. Kant argues instead that we are creatures who live our desires and our needs as do all other animals, and yet we are creatures who have the possibility – and let me emphasise the word possibility, because it can neither be theoretically demonstrated nor theoretically refuted as a fact about ourselves – that we can act other to the mechanics of those desires and needs.

Kant is often accused of being a prude because Kantian negative freedom is mistakenly identified with not doing what you want to do. But that is not the case. For Kant, a human life is purposive, and when we take ourselves seriously as creatures who can set ends for themselves as long-term directions in life, we should also find ourselves capable of coordinating our purposes and ends with the ends of others. Indeed, the idea of humanity,
famously presented as a formulation of the categorical imperative, is inseparable from the possibility that each of us can project our ends as a reasonable creature who can promote a community in which ‘mine’ harmonises with ‘yours’ in the Kingdom of Ends.4

Negative freedom can entail limiting our own immediate desires through our attempt to live a purposive life, but it is not simply antagonistic to needs or desires. As finite beings we must make choices. We cannot be everywhere at once, and if we did not make choices we would not be able to represent them as part of our own life project. In this specific sense we have to exercise our prudence. Positive freedom for a creature that must live in a finite body, bound by the laws of nature, must proceed through a notion of autonomy that reconciles law and freedom. Indeed, positive freedom for Kant is causality through moral action, in which each of us potentially creates a new beginning as a moral person and with other persons, and aspires to create a new social and political order guided by the Kingdom of Ends.

For Kant, autonomy is not at all reducible to being left alone by others – not even in a more affirmative sense of aloneness as self-determination. This is precisely because for Kant freedom must be law-like, which means we exercise our autonomy not when we are strictly self-determining – true self-determination is beyond the reach of a finite creature for Kant – but rather when we represent our ends as our own and not merely determined by the pulls and tugs of the day to day world; in the strongest sense, we can only hope to attain freedom by imposing a law upon ourselves which allows us to represent ourselves as the source of a different ethical causality, a new moral beginning. Negative and positive freedom together, as Kant understands them, should be taken to mean that it is human beings who set value on ends; we make them ours in that we project an ‘I’ as a necessary postulate of the very practical reason to which we must return to justify our ends.

To understand freedom and autonomy in this way is extremely significant for the law of the new South Africa, as the judgments of the emeritus justice, Justice Ackermann, demonstrate. First, as I have already suggested, the Kantian notion of positive freedom or autonomy gives us a defence of equality that does not depend on human persons being materially alike – a problem that has plagued some theorists in the Anglo-American academy. Human beings have equal worth because we all have the

4 The concept of every rational being as one who must regard himself as giving universal law through all the maxims of his will, so as to appraise himself and his actions from this point of view, leads to a very fruitful concept dependent upon it, namely that of a kingdom of ends. By a kingdom I understand a systematic union of various rational beings through common laws. Now since laws determine ends in terms of their universal validity, if we abstract from the personal differences of rational beings as well as from all the content of their private ends we shall be able to think of a whole of all ends in systematic connection (a whole both of rational beings as ends in themselves and of the ends of his own that each may set himself), that is, a kingdom of ends, which is possible in accordance with the above principles’ (Groundwork for the metaphysics of morals (1997) 41).
possibility of guiding our actions through practical reason. We should, as a moral mandate – and in the case of Justice Ackermann’s ‘New South Africa’ also as a constitutional mandate – regard all other human beings through the representation of that possibility. Thus, Kant introduces the idea of horizontal thinking, by which I mean the recognition that all creatures defined by this possibility have equal worth. Therefore, any return to hierarchies that deny to any human being that he or she partakes in this possibility must be rejected.

As an ideal attribution, dignity can be violated even as it remains the basis of our moral worth, because as creatures that have at least the possibility of moral action we might be able to bring about a new beginning. That is, dignity as an ideal attribution of our sameness before this possibility can be violated, but not lost. Here we see the difference between sameness and likeness for Kant. We are the same before this possibility even if as a matter of actual life many, if not all of us, give ourselves long (sometimes endless) moral holidays. Nor does it matter for a defence of our intrinsic equal worth that we are actually different in our physical selves. Kant defends a view of autonomy rooted in how we might be: In this sense he does not offer a moral or ethical ontology of who we are rooted in any theory of human nature. Indeed, to the degree that Kant speculated in the anthropology of his time, he had a rather sour view of human nature.

For the former Justice Ackermann, the moral categorical imperative for the new South Africa is precisely the recognition of this sameness, which, of course, is a complete moral inversion of the dreadful history of apartheid. For the former Justice Ackermann, the constitutional imperative demands that never again will dignity be violated, and that its respect requires that dignity not only be treated as a right (as it must be under section 10 of the Constitution) but also as an ideal that informs how the other rights in the Bill of Rights are to be interpreted. On this understanding, when a justice uses dignity as a constitutional imperative, that justice is not locked into a kind of formalistic reasoning that says we respect the Constitution because it is the Constitution – as some critics of the dignity jurisprudence have suggested. For Ackermann, it is dignity that is the grundnorm of the entire Constitution.

Again, a grundnorm is the moral or ethical principle or ideal that undergirds the moral as well as the legal order of society. We need to note that the distinction between moral and legal rights is only made in Anglo-American jurisprudence, while for Kant legal rights are moral rights. The meaningful distinction in Kant is between the realm of internal freedom – morality – and the realm of external freedom, law. More on that distinction shortly.

In the case of South Africa, this understanding of dignity can yield an interpretation of section 10 of the South African Constitution that separates
the two clauses of section 10. If everyone has dignity as an ideal attribution, then it is that ideal attribution that is recognised as a right. Dignity is what Justice Ackermann establishes as the *grundnorm*, following Kelsen’s definition, and he thus places dignity at the heart of the substantive revolution of South Africa. For Kelsen, two kinds of revolution are possible. The first is the more familiar kind of revolution, a full revolution. A full revolution in South Africa would have obliterated the entire preceding legal system, which would have included firing all the judges and declaring all former law invalid. A substantive revolution, on the other hand, is when the new order proceeds legally even as the old order validates the creation of a new government – and yet it still turns the world upside down by creating a new objective, normative legal order that is based on an ethical principle that negates the ethical acceptability of the old order. Hence, the word revolution applies. Under Justice Ackermann’s dignity jurisprudence, the South African Constitution is the law of a substantive revolution. The newness of this process is that it is one of a very few substantive revolutions to take place in world history (depending on how one reads some of the new dispensations in South America). Thus, if the Constitution ceased striving to embody the ideals of an objective normative order worthy of a substantive revolution, it would no longer be worthy of respect as the supreme law of the land of the new South Africa. This is not what the letter of the Constitution says, but it is the very spirit of the constitutional mandate that Justice Ackermann seeks to express in his judgments concerning the moral and constitutional significance of dignity.

But how is the social bond that undergirds Kant’s notion of law as the realm of external freedom represented? The relationship between the realm of internal freedom (of morality) and the realm of external freedom (of right or *Recht*) has long been debated in Kantian scholarship. As Allen Wood explains:

> It is unclear whether Kant holds that the two moral legislations of right and ethics are derived from a common principle such as the supreme principles of morality that receives a three-fold formulation in the groundwork. It speaks in favour of a common or unified basis that Kant regards duties of right as categorical imperatives and also that he grounds the single innate right possessed by persons (the right to freedom) on the humanity of persons (hence apparently on the second main formula of the principle of morality, the formula of humanity as an end in itself)[…] But it speaks against a unified basis of right and ethics that the principle of right is described as analytic, whereas the principle of morality is synthetic, and also that the principle of right itself does not actually command us to do anything, but merely tells us what it takes for an action to have the status of a ‘right’ action within the system of right.7

---

5 Sec 10 of the Constitution of the Republic of South Africa, 1996 reads, ‘Everyone has inherent dignity and the right to have their dignity respected and protected.’

6 Kelsen (p 3 above) 117.

But clearly, and despite Kant’s waffling, there must be a connection between the two. If there were no connection, there would be no moral ground for the realm of external freedom in which we coordinate our ends with one another. Kant’s hypothetical experiment in the imagination, in which we configure the conditions in which human beings could aspire to the great ideal of the Kingdom of Ends, turns on the possibility that as creatures of practical reason we can harmonise our interests. For Kant, we represent the realm of external freedom through a hypothetical experiment of the imagination in which we configure the conditions of a social contract rooted in the respect for all other human beings. Under this experiment in the imagination, we imagine the conditions in which individuals are given the greatest possible space for freedom, as long as it can be harmonised with the freedom of all others. The social contract imagines us as moral beings that can exercise their practical reason and potentially guide their actions in accordance with its mandates. In his essay ‘On the common saying: “This may be true in theory but it does not apply in practice”’, Kant argues that the Hobbesian social contract will always falter precisely because the basis for that contract is not rooted in our potential to act morally.\(^8\) Hobbes’s social contract imagines human beings as forsaking their natural liberty and yielding to the coercive power of the sovereign and of the positive law only out of the drive for security and the protection of expectation. Thus, simply put, if the only basis for abiding by a legal system is fear and security with no moral reason for obligation, there will always be a reason for opting out of the social contract.

3 Criticisms of the Kantian defence of dignity

It is impossible to rehearse the rich critical literature on Kant in my short talk today. But let me paint some of the most well-known critiques of Kantian dignity with a broad brush. I will focus particularly on those criticisms that will help us bring into focus the similarities and differences between ubuntu and the Kantian notion of dignity.

The first criticism is that since our dignity lies in our potential to abide by the dictates of pure practical reason, and most of us do not do this most of the time, dignity so defended is out of touch with the reality of human nature. Secondly – and many feminists have passionately asserted this criticism – what of human beings who for some reason cannot act rationally, even as a potential? Do they not have dignity? Even if dignity is an ideal attribution it turns on a potential or possibility that some people do not have because they lost it at birth or have yielded to the physical realities of old age or illness that undermine rational capacity. Dignity so

\(^8\) I Kant ‘On the relationship of theory to practice in political right’ in HS Reiss (ed) Kant’s political writings (2001).
conceived, then, is too narrow in its reach. It fails to assert the dignity of all human beings.9

Next, the Kantian notion of autonomy is sometimes critiqued as too individualistic. Yet this criticism fails to grasp how human beings belong together in primordial ways. Critics claim that even the hypothetical imagined social contract, imagining us as capable of acting morally and building a new world together, still begins with individuals. A version of this criticism is that human beings are never truly self-determining and thus autonomy is not a possible representation of ourselves as moral creatures: It is a fantasy, a dangerous one that can end up justifying self-righteous moralism. Feminists have emphasised another aspect of this criticism, which is that autonomy denies human fragility and our need for care and support.10 This feminist emphasis on human fragility argues that Kantian autonomy is a male myth that bolsters a sexist view of the human. Another related criticism is that Kant grounds our dignity in a mystical noumenal otherness outside our actual phenomenal lives. Thus, our dignity seems to pit itself against both individual happiness and a collective promise of the common good.

To be fair to Kant, some of these criticisms are based on misreadings of Kant’s critical idealism. As I have already suggested, autonomy is not self-determination for Kant but rather our determination by a law, albeit a law that we lay down to ourselves. Thus, Kant is not an individualist, at least not in the common English understanding of the word. To the degree that we are creatures of intrinsic worth, we have autonomy because of a sameness that we share. Dignity does not turn on individual characteristics or even some essential idea of our singularity. Autonomy is not a fact of our individuality. Nor does Kant argue that we can know for sure that we are acting rightly. As phenomenal creatures we can always get it wrong even as we try to abide by the moral law. In the Critique of practical reason, Kant himself powerfully argues against the dangers of self-righteousness. This said, Kant does argue that our dignity is inextricably associated with our capacity for reason. The social bond, even if it aspires to our acting together under the guidance of the great ideal of the Kingdom of Ends, still begins with imagined individuals.

4 The alliance between African humanism and Kantian ethics

As we will see, ubuntu does not defend dignity through our capacity for reason and does not think that the social bond is an experiment that begins with imagined moral individuals. And yet, there is an alliance of a Kantian notion of dignity with the intellectual heritage of African Humanism. The connection I am drawing with African humanism should not be taken to mean that there is

---

10 As above.
nothing unique about *ubuntu* or that African humanism is not a complex tradition differentiated both by time and space. But since this intellectual heritage is often ignored or disparaged, it is important to remind ourselves of the important role it has played in Afro-modernity, including its role in the struggles for national independence and the re-emergence of an ethical view of socialism that was deployed in economic development. Africa has a complex history of attempts to realise an ethical ideal of socialism, and I in no way want to ignore that history, nor do I want to simplify it. Fortunately, much deeper work is now being undertaken to study what was and what was not achieved, for example, in Tanzania under the leadership of Jules Nyerere. My emphasis in this essay is on the intellectual heritage to the degree that there are important links with *ubuntu*. But let us now turn again to the alliance of Kantianism with African humanism.

To some degree this alliance should not be surprising because Kant, and in particular Kant’s notion of dignity, has played a major role in the work of European thinkers who also argue that socialism is an ethical ideal and not a stage in a Marxist science of development and progress. Kwame Nkrumah, for example, argues that his philosophy of consciencism was an attempt to find an alternative justification for Kantian ethics, and particularly dignity, consistent with the African conscience. To quote Nkrumah:

> If ethical principles are founded on egalitarianism, they must be objective. If ethical principles arise from an egalitarian idea of the nature of man, they must be generalisable, for according to such an idea man is basically one in the sense defined. It is to this non-differential generalisation that the expression is given in the command to treat each man as an end in himself, and not merely as a means. That is, philosophical consciencism, though it has the same ethics as Kant, differs from Kant in founding ethics on a philosophical idea of the nature of man. This is what Kant describes as an ethics based on anthropology. By anthropology Kant means any study of the nature of man, and he forbids ethics to be based in such a study.11

More commonly, the dignity of human beings is connected to a phenomenology of African social life, including the struggle against colonialism. This grounding in other African thinkers, as we will soon see, if it is in the phenomenological sense African, is still ethically justified as universal, even without Nkrumah’s attempt to develop a unique African notion of materialism that is to serve as the basis for an egalitarian ethic that would re-ground Kantian dignity.

Kenneth Kaunda, the first president of Zambia, also argues that African humanism is a different approach to Kantian ethics, and yet one that also insists on dignity. Let me quote Kaunda’s phenomenology of the

---

African extended family as it helps us to think differently about the basis for ethics:

The extended family system constitutes a social security scheme, which has the advantage of following the natural pattern of personal relationships rather than being the responsibility of an institution. It also provides for richness in knowledge and experiences for those fortunate enough to be part of it. Granted, I have been describing the character of small-scale societies and it could be argued that such a system would not work where hundreds of thousands of people are gathered together in cities or towns. But the attitudes to human beings, which I have set out, are not solely a function of social organisation. They are now part of the African psychology. I am deeply concerned with this high valuation of Man and respect for human dignity, which is our legacy. Our tradition should not be lost in the new Africa, however ‘modern’ and ‘advanced’ in a Western sense the new nations of Africa will become. We are fiercely determined that this humanism will not be obscured. African humanism has always been Man centered. We intend that it remain so.12

Julius Nyerere echoes Kaunda in his summation of the Arusha Declaration; the Arusha Declaration is also a commitment to a particular quality of life. It is based on the assumption of human equality, on the belief that it is wrong for one man to dominate or exploit another and on the knowledge that every individual hopes to live in a society as a free man to lead a decent life in conditions of peace with his neighbors. The Declaration, in other words, is Man-centered. Inherent in the Arusha Declaration, therefore, is a rejection of the concept of national grandeur, as distinct from the well-being of its citizens, and a rejection too of material wealth for its own sake. It is a commitment to the belief that there are more important things in life than amassing riches, and that if the pursuit of wealth clashes with things like human dignity and social equality, then the latter will be given priority.13

Kaunda further argues that there are three key social virtues that are crucial to African humanism: mutuality, acceptance and inclusiveness. The intrinsic worth of a human being, which justifies an egalitarian ethic, is not based, as it is in Kant, in that being’s potential to act in accordance with the dictates of reason. Nor is the worth of a human being derived from his or her actual achievements. As Kaunda emphasises:

The success-failure complex seems to be a disease of the age of individualism – the result of a society conditioned by the diploma, the examination, and the selection procedure. In the best tribal society people were valued not for what they could achieve but because they were there. Their contribution, however limited to the material welfare of the village was acceptable, but it was their presence, not their achievement which was appreciated.14

13 J Nyerere ‘The purpose is man’ in J Nyerere Uhuru na Ujamaa (1968) 316.
14 Nyerere (n 13 above) 26.
A human being has dignity because she is a unique being born into the human community, raised and supported there. For Kaunda, the keystone for judging any society is whether the older people are respected. ‘The fact that old people can no longer work, or are not as alert as they used to be, or even have developed the handicaps of senility in no way affects our regard for them. We cannot do enough to repay them for all they have done for us. They are embodied wisdom; living symbols of our continuity with the past.’

5 Ubuntu and African humanism

We may seem to be far away from ubuntu, but indeed we are not, precisely because ubuntu also emphasises the virtues of mutuality, inclusiveness and acceptance. In like manner, respect for dignity is rooted in singularity and not in our capacity for rationality. Emeritus Justice Yvonne Mokgoro, has explained ubuntu as follows:

Generally ubuntu translates as humaneness. In its most fundamental sense, it translates as personhood and morality. Metaphorically, it expresses itself in umuntu ngumuntu ngabantu, describing the significance of group solidarity on survival issues so central to the survival of communities. While it envelops the key values of group solidarity, compassion, respect, human dignity, conformity to basic norms and collective unity, in its fundamental sense it denotes humanity and morality. Its spirit emphasises respect for human dignity, marking a shift from confrontation to conciliation.

Ubuntu so understood is both a principle of transcendence for the individual as each one of us transcends our biological distinctiveness that marks us as unique from the beginning of life as we struggle to become a person. The burying of the umbilical cord marks the biological distinctiveness of each person. We are distinct from that moment of the cutting of the umbilical cord, with an origin shared by no other. But we only become a unique person through a struggle to define who we are, and we only do this with the support of others who help us in our effort at self-definition. Thus the law of transcendence for the individual, ubuntu, is also the law of the social bond. And why is this the case?

In ubuntu individuals are intertwined in a world of ethical relations and obligations from the time they are born. This inscription by the other is not simply reduced to a social fact. We come into the world obligated to others, and in turn these others are obligated to us, to the individual. Symbolically this is demonstrated in the burying of the umbilical cord. Others do that for us and create the origin from which we began our unique journey. But what is marked in this ritual is our unique-ness. Thus, it is a profound misunderstanding of ubuntu to confuse it with simple-minded communi-

---

15 As above.
tarianism. It is only through the support of others that we are to realise true individuality and rise above and beyond our biological distinctiveness into a person whose singularity is inseparable from the journey to moral and ethical development. Individuation is an achievement that involves struggle and the acceptance of rites of passage. Ifeanyi Menkiti captures this relationship between biological distinctiveness and personhood in the following passage:

In the stated journey of the individual toward personhood, let it therefore be noted that the community plays a vital role both as catalyst and as prescriber of norms. The idea is that in order to transform what is initially a biological given into full personhood, the community of necessity has to step in, since the individual himself or herself, cannot carry through the transformation unassisted. But what are the implications of this idea of a biologically given organism having first to go through a process of social and ritual transformation, so as to attain the full complement of excellences seen as definitive of the person?

One conclusion appears inevitable, and it is to the effect that personhood is the sort of thing which has to be achieved, the sort of thing at which individuals could fail. I suppose that another way of putting the matter is to say that the approach to persons in traditional thought is generally speaking maximal, a more exacting approach insofar as it reaches for something beyond such minimalist requirements as the presence of consciousness, memory, will, soul, rationality or mental function. The project of being or becoming persons, it is believed, is truly a serious project that stretches beyond the raw capacities of the isolated individual, and it is a project laden with the possibility of triumph, but also of failure. 17

If the community is committed to individuation and the achievement of a unique destiny for each person, the person in turn is obligated to enhance the community that supports him or her, not simply as an abstract duty that is correlated with a right, but as a form of participation that allows the community to strive for fidelity to what Masalo has called participatory difference. 18 For Masalo, this participatory difference recognises that each one of us is different, but also that each one of us is called upon to make a difference by contributing to the creation and sustenance of a humane ethical community. We can understand, then, that our ethical relationship to others is inseparable from how we are both embedded in and supported by a community that is not outside of us, something over ‘there’, but is inscribed in us. The inscription of the other also calls the individual out of himself or herself. Ubuntu is in this sense a call for transcendence. The individual is called back towards the ancestors, forwards toward the community, and further towards relations of mutual support for the potential of each one of us.

18 D Masalo ‘Western and African communitarianism’ in Wiredu (n 17 above) 496.
We have seen, however, that this does not imply any simple notion of communitarianism or social cohesion. Each one of us is called to become her own person and make a difference so as to realise the ethical quality of humanness, and to support such a quality of humanness in others. Although others support me, and an ethical action is by definition ethical because it is an action in relationship to another human being, it is still up to me to realise my own personal destiny and to become a person in the ethical and moral sense of the word. Thus, it is humanity, as Justice Mokgoro emphasises in her own definition of *ubuntu*, and not just my community that is at stake in my actions. If I relate to another person in a manner that lives up to *ubuntu*, then there is at least an ethical relationship between the two of us. The concept of a person in *ubuntu* is an ethical concept. A self-regarding or self-interested human being is one that has not only fallen away from her sociality with others; she has lost touch with her humanity. One crucial aspect of doing justice to such a person is that we who are participating in an ethical community help that individual get back in touch with himself or herself. Thus, cohesion and harmony are not the ultimate good because they must always be submitted to the doing of justice. As Murungi argues:

Certainly in Africa, but not only in Africa, personhood is social. African jurisprudence is part of African social anthropology. Social cohesion is an essential element of African jurisprudence. Areas such as criminology and penology, law of inheritance and land law, for example focus on the preservation and promotion of cohesion. This cohesion is a cohesion tempered by justice. Justice defines a human being as a human being. Thus, injustice in Africa is not simply a matter of an individual breaking a law that is imposed on him or her by other individuals, or by a collection of individuals who act in the name of the state.  

It is a violation of the individual’s duty to himself or herself, a violation of the individual to be himself or herself – the duty to be a social human being.

Thus, this specifically ethical manner in which personhood is conceived has important implications for how the social bond that underlies any notion of law is defined. This means of course that this idea of personhood, as Murungi points out, underlies the central concepts of African jurisprudence, as these in turn set the meaning of law. Again, to quote Murungi:

Each path of jurisprudence represents an attempt by human beings to tell a story about being human. Unless one discounts the humanity of others, one must admit that one has something in common with all other human beings. To discount what one has in common with other human beings is to discount oneself as a human being. What is essential in law is what secures human

---

beings in their being. The pursuit and the preservation of what is human and what is implicated by being human are what, in a particular understanding, is signified by African jurisprudence. Being African is a sign of being African and a sign of being human. In this signature lies not only what is essential about African jurisprudence, but also what is essential about the Africanness of African jurisprudence. To learn how to decipher it, which in a sense implies learning how to decipher oneself paves the way to a genuine understanding.  

In a profound sense then, African jurisprudence articulates the principles of African humanism. Ubuntu can and should be grasped as part of this other jurisprudence and view of law. It is indeed a different story of law and justice than the one imagined and told in the West. We will return to that difference when we discuss the contrasts and similarities between Kantian dignity and ubuntu.

For now I want to discuss Justice Mokgoro’s judgment in the Khosa case, a judgment in which she does not actually use the word ubuntu, but one that is infused with ubuntu thinking. The case raised the issue of whether non-citizens who were permanent residents would be eligible for certain social welfare grants as long-standing members of the community. Justice Mokgoro’s judgment granted such eligibility, but her justification for so doing is what is important for our purposes. Remember Kaunda emphasised three main virtues of African humanism: mutuality, acceptance and inclusion. All three entered into Justice Mokgoro’s reasoning in the Khosa case. First, the overriding idea of African jurisprudence is reflected in the judgment. What would it mean to deny non-citizens social welfare grants if ‘what is essential to law is what secures human beings in their being?’ Further, Justice Mokgoro emphasises mutuality. The people who are seeking grants are members of a community, and mutuality here means that we must respect their contributions to the community, but more profoundly that they are part of us and we are part of them.

This turns us back to an earlier point in this essay in which I emphasised that the community is not something out there, but instead signals how we are inscribed in each other. Just as non-citizens have obligations to the community, so the community, including the state, has obligations to them. Secondly, a humane community is both accepting and inclusive. To refuse to grant them eligibility would render them outsiders to whom we do not owe relations of mutuality. In a sense then, to deny them social grants would violate their dignity precisely because the mark of outsider is not only to push another out of the community, but also outside what is owed to another human being as a human being. They would be pushed outside of the human world of mutual relations. Throughout, Justice Mokgoro emphasises actual

20 Murungi (n 19 above) 525 - 526.
21 For those who wish to read a review of the cases both in the Constitutional Court as well as the lower court, see Cornell (n 2 above).
belonging and interconnectedness, which underscores how mutuality, acceptance and inclusiveness reinforce one another. If we ignore the needs of those who have become part of us, then we are at risk of losing our own humanity. The emphasis, then, is not so much on dignity as an individual right, although Justice Mokgoro does appeal to dignity, but dignity understood as the ethical core of relations of mutuality.

6 Criticisms of ubuntu

Again, we will paint the criticisms of ubuntu in broad brushstrokes. First, there is the criticism that ubuntu is a communal ethic that denies the importance of individual autonomy. It emphasises obligations to the group at the expense of individual rights. Some critics read umuntu ngumuntu ngabantu to mean that a person is only a person through other people. This reading promotes the criticism that there is no place for individuality and creativity in an ubuntu ethic. Secondly, there is a criticism that ubuntu is inherently conservative. Worse yet, its appeal to cohesion privileges, dangerous hierarchies, corrupt tribal authorities over the people they supposedly lead, men over women, etc. Thus, ubuntu is both authoritarian and patriarchal and should be mistrusted as such. Thirdly, even if ubuntu was once an important value in the struggle for liberation and indeed, for survival in horrific circumstances, it is dying out under the force of advanced capitalism. Thus, what remains of ubuntu is a shadow of what it used to be and as a result, it easily falls prey to manipulation, not just by corrupt traditional authorities but also by the African National Congress leaders who want to maintain a compliant workforce in the face of the demands of neo-liberal capitalism. Given these three criticisms, ubuntu is rejected for being backward looking and not forward looking. Thus it is denied that ubuntu has the same aspirational edge as dignity as an ideal.

Before turning to the contrast with Kantian dignity, let us just remember that some of these criticisms involve a fundamental misunderstanding of ubuntu. Often, critics of ubuntu make the mistake of reducing ubuntu to an ethical ontology of a purportedly shared world. What is missed in this criticism is precisely the activism in participatory difference. This activism is inherent in the ethical demand to bring about a humane world. Ubuntu clearly has an aspirational and ideal edge – there is no end to a struggle to bring about a humane world, and who makes a difference in it. More has given us a rich definition of ubuntu which captures the specificity of how ubuntu has an aspirational edge. To quote More:

In one sense ubuntu is a philosophical concept forming the basis of relationships, especially ethical behaviour. In another sense, it is a traditional politico-ideological concept referring to socio-political action. As a moral or ethical concept, it is a point of view according to which moral practices are founded exclusively on consideration and enhancement of human well-being; a preoccupation with the human. It enjoins that what is morally good, dignity, respect, contentment, and prosperity to others, self and the community at
large. *Ubuntu* is a demand for respect for persons no matter what their circumstances may be.

In its politico-ideological sense it is a principle for all forms of social or political relationships. It enjoins and makes for peace and social harmony by encouraging the practice of sharing in all forms of communal existence.\(^{22}\)

*Ubuntu* as an ethical, as well as a politico-ideological concept, is always integral to a social bond. In a profound sense, *ubuntu* encapsulates the moral relations demanded by human beings who must live together. As we have seen, it implies a fundamental moralisation of social relations, and what never changes is that society is inherent moral. But the actual relationships *ubuntu* calls us to must change since *ubuntu* is inseparable from a relationship between human beings that is always present, yet it is also connected to how we are always changing in those relationships and our needs are changing with them. The aspirational aspect of *ubuntu* is that we must strive together to achieve the public good and a shared world. It is *ubuntu*'s embeddedness in our social reality that makes it a transformative ethic at its core. It is the moralisation of social relationships that inevitably demands the continuous transformation of any society. It would have been absurd 500 years ago to put forward the argument that *ubuntu* demands access to electricity, but it would not be absurd to do so now. Indeed, electricity is integral to securing a human life in modern society.

In like manner, *ubuntu* has no pre-given position on the equality of women. Certainly there is sexism in African thinking as there is in African social and political life. *Ubuntu* has been used to justify sexism in the name of tradition, and it has been deployed on the other side to fight it. That it can be used on different sides of a struggle over what is right does not mean that *ubuntu* can mean anything. It does mean that it is embedded in a complex and changing world, so that what was believed to be right for women hundreds of years ago has been profoundly challenged by women on the ground. Again, to turn to Murungi’s definition that ‘law secures human beings in their being’, law in the sense of doing justice does not separate between civil and socio-economic rights. Both are necessary to secure and protect our humanity in the moral sense that is echoed in Emeritus Justice Mokgoro’s *Xhosa* judgment and in the writings of the African philosophers – all of whom represent different aspects of African humanism.

Some of the criticisms represent a misunderstanding of *ubuntu*, as we have seen throughout my discussion of *ubuntu*. *Ubuntu* does not undermine the concept of personhood; indeed, it demands the thinking of the ethical conditions of individuation. Yet it does not embrace individualism and rejects the notion of self-determination. We will return

\(^{22}\) P More ‘Philosophy in South Africa under and after apartheid’ in Wiredu (n 17 above) 149, 156 - 157.
to the significance of this mis-reading of Kant shortly for the history and the future of the ideal of dignity. As I have already argued, Kant also does not reduce autonomy to self-determination. Whether or not ubuntu has been undermined by modernity, apartheid, and capitalism is an empirical question and must be addressed as such. At least in the townships of the Western Cape, the Ubuntu Township Project has undertaken to investigate whether an ubuntu ethic is alive, and what are the forces that either support it or undermine it. Such empirical work can also help us grapple with the undoubted reality that some leaders in South Africa have manipulated ubuntu for their own purposes. Again, we will return to these important criticisms of ubuntu.

7 Similarities with Kant

For now let me turn to ubuntu’s similarities with Kantianism. We have already seen that there is a historical alliance between Kant and African humanism. I want to emphasise two crucial aspects of that alliance.

First, both draw a close connection between freedom and morality, freedom and obligation, freedom and necessity. One is not free outside of the ethical, but through it. A human being cannot be free by doing whatever he or she wants. We are constrained by the laws of nature. Our freedom is expressed in our laying down a particular kind of law unto ourselves. It is a particular kind of law, but it is a law. When we live up to the obligations of the moral law we are also representing ourselves as free. What has come to be known as the reciprocity thesis, that freedom and the obligations imposed upon us by the moral law are two sides of the same coin, should show us how far we are, in Kant, from the identification of freedom with self-determination. In his later writing, Kant also argues that it is not only as a moral personality that we have dignity but also through our humanity.23 Kant underscores that even when we act in our self-interest we do so through reason, by making a desire a maxim that we choose to follow. In Religion within the boundaries of mere reason, Kant holds that reason guides us in our humanity as well as in our moral personality, a distinction he did not make in his earlier work.

But that said, it is through our capacity for practical reason and our potential to be dictated by the demands of the moral law that humanity itself becomes a moral ideal. And here we find the second basis for the alliance with African humanism. Humanity matters because it is an ethical ideal. Surely there are different notions of the meaning of the ethical ideal of humanity, but broadly construed they share the insistence that personhood and morality are inseparable. Hence, both Kantianism and African humanism strongly promote human dignity. It is important to note here that the misreading of Kantian autonomy that identifies it with

23 I Kant Religion within the boundaries of mere reason (2004) 50.
self-determination has had serious consequences for the Western discourse on autonomy. As Berkowitz has accurately argued:

The problem that plagues all of these modern invocations of dignity from Dworkin to Habermas is that they understand dignity as self-determination and then stumble upon the contradiction in self-determination. Self-determination – the right to personal autonomy that underlies human and civil rights – is frequently at odds with self-determination – the democratic right of citizens to collectively govern themselves. Indeed, the tension between individual and collective self-determination is at the root of many of the great human rights tragedies of the modern era.24

For Arendt, this dilemma, which has haunted human rights discourse, cannot easily be theoretically or practically overcome since it lies at the heart of modern liberalism. And so she calls for a new political principle, a new law on earth.25

As we have seen, Kant himself did not identify autonomy with self-determination. Yet there is little doubt that most of his modern interpreters have done so. Can we still rely on Kant as the basis for dignity? Or do we need another discourse? Even under the most sophisticated interpretation of Kant, our dignity lies in our capacity for reason. For a thinker like Arendt that capacity has carried little force against the horrors of the 20th century. Kant of course did not base his idea of dignity in how we actually act but in how we should act, and it is always a possibility for us that we might act rightly if we abided by the moral law. But is that enough for a new law on earth?

8 Differences between Kantianism and ubuntu

There are several important differences that we need to emphasise between Kantianism and ubuntu. First, as we have seen, dignity in ubuntu thinking is not rooted in reason because of an ethical concern shared with many feminists that this would deny dignity to too many human beings. Thus, such a ground for dignity runs afoot of the virtues of inclusiveness and acceptance. Instead, dignity is rooted first and foremost in our singularity and uniqueness, and at the same time in our embeddedness as part of a human community.

As we have also seen, the social bond is conceived differently, and thus, although both traditions moralise law – and would find it equally incomprehensible to separate law and morality in the Anglo-American sense – they rely on different conceptions of moral personhood to do so. Even the great Kantian hypothetical experiment in the imagination, in

---

25 H Arendt The origins of totalitarianism (1973) ix.
which we configure the conditions of the social contract rooted in the respect for all other human beings, still begins with imagined moral individuals. It is still individuals who agree to accept some form of coercion, even if rooted in Kant’s basic understanding of right, which is that individuals are allowed the greatest possible space for their freedom as long as it can be harmonised with the freedom of all others in the social contract. If freedom is inseparable from morality then at least as a possibility, we can represent ourselves as free human beings who can harmonise their interests in the Kingdom of Ends.

Ubuntu, alternatively, does not conceive of the social bond through an imagined social contract undertaken by imagined moral individuals. We are born into a social bond but it is not as if the social is something outside the individual. It is the network of relationships in and through which we are formed and whose formation is ultimately our responsibility. There is a flow, back and forth, between the individual and others as he or she undertakes the struggle to become a person, always conceived ethically, which is difficult to think of in Western philosophy. It is certainly not impossible to think of such a back and forth, but it is difficult because the social is sedimented as something outside the individual. Of course, if we add Kant to Hegel and Marx, we begin to see possibilities for such a dialectic. Am I suggesting that ubuntu thinking and African humanism are better resources on which to ground the new law on earth, the new political principle to which Arendt calls us? I am certainly arguing that a reinvigorated defence of dignity, whether as the Grundnorm of the South African Constitution or as a new universal political and ethical principle must bring this intellectual heritage into the debate as an equal partner. Such a dialogue is long overdue, and necessary for the continuing defence of human dignity in South Africa and the rest of the world.
If I am not for myself, then who will be for me?
If I am not for others, then who am I?
If not now, when?
Hillel

Haeba ke sa ithokomele, ke mang ea tla nthokomela? Haeba ke sa hlokomele ba bang ba heso, na ke botho? Haeba re sa hlokomelana ha joale, re tla hlokomelana neng?
Anonymous

History doesn’t repeat itself, but it often rhymes.
Mark Twain

In the decade or so in which Professor Cornell has engaged South Africa’s jurisprudence, her name has become synonymous with academic discourse about and around the values of dignity and ubuntu. As colleagues and collaborators, it is often hard to know where Drucilla Cornell’s thoughts on these subjects end and one’s own ruminations begin. What follows then is an amplification of, or a riff upon, Professor Cornell’s lead essay: ‘Is there a difference that makes a difference?’ Contestation is not on the cards.

* Justice Emeritus, Constitutional Court of South Africa; Professor Extraordinaire, Faculty of Law, University of Pretoria.
** Professor and Elizabeth Bradley Chair of Ethics, Governance and Sustainable Development, University of the Witwatersrand; Co-Director, South African Institute for Advanced Constitutional, Public, Human Rights and International Law; Editor-in-Chief, Constitutional law of South Africa.

1 We have both written at length on dignity and ubuntu, and in many cases in conjunction with, or with support from, Professor Cornell. See, eg, Y Mokgoro ‘Ubuntu and the law in South Africa’ (1998) 4 Buffalo Human Rights LR 15; D Cornell & Y Mokgoro (eds) ubuntu in the law of South Africa (2012); S Woolman ‘Dignity’ in S Woolman & M Bishop (eds) Constitutional law of South Africa (2005) (2nd ed); ch 36; D Cornell, S Woolman, M Bishop, J Brickhill & S Fuller (eds) The dignity jurisprudence of the Constitutional Court of South Africa (2012) (forthcoming).
Chapter 10

But emotion is. Adjudication and academic scholarship in the social sciences – as much as any other form of judgment – requires ‘emotional intelligence’ and a kind of connected spectatorship based upon mutual respect, care, trust, loyalty and respect. As Professor Cornell insists throughout her immense and influential body of work, law cannot be reduced to a benighted Hartian set of first order rules of recognition and the second order rules that flow from them. Cornell’s persistent emphasis on law as far more than a body of rules functions as a regular reminder that our ‘formal’ training (at institutions as physically far apart as the University of Bophuthatswana or Columbia Law School, but ontologically proximate with respect to the Hartian indoctrination into legal doctrine that we received) was missing something critical to the creation of a just legal order. No matter how much law strives to have a formal existence, messy human beings living out different ways of being in the world must learn to accommodate one another when it comes to determining the meaning of the basic law and honouring the constraints that the text invariably imposes. Following Botha and Tribe, we write neither ‘for those who feel confident that canons of appropriate constitutional construction may be derived from some neutral source’ nor ‘for those who have convinced themselves that anything goes as long as it helps end what they see as injustice’. When we write about the relationship between ubuntu and the Constitution, we provide substantive grounds (expressly present in the interim Constitution and implicit in the final Constitution) for how we believe judges on a South African bench ought to read the text. Put slightly differently, though ubuntu may shadow Western notions of dignity (drawn

---

2 See S Ellmann ‘Marking the path of the law’ (2009) 2 Constitutional Court Review 97 (Ellmann contends that emotional attachment, as opposed to detachment, plays a greater role in adjudication than even legal realists tend to posit and that the role of emotional attachment in the law ought, as a result, to be given greater recognition in the development of the law school curriculum (primarily in the form of an increased emphasis on clinical courses).)

3 See A Smith The theory of moral sentiments (1760). See also A Sen The idea of justice (2010); A Sen Development as freedom (1999).

4 See D Cornell The imaginary domain (1995); D Cornell At the heart of freedom (1998); D Cornell Just cause: Freedom, identity and rights (2000); D Cornell Defending ideals: War, democracy, and political struggles (2004); D Cornell Moral images of freedom (2008); D Cornell ‘A call for a nuanced constitutional jurisprudence: Ubuntu, dignity and reconciliation’ (2004) 19 SAPL 666 667 (‘[I]f we give Kantian dignity its broadest meaning, it is not associated with our actual freedom but with the postulation of ourselves as beings who not only can, but must, confront ... ethical decisions, and in making those decisions ... give value to our world.’). For an insightful overview of Cornell’s oeuvre, see K van Marle ‘“No last word”: Reflections on the imaginary domain, dignity and intrinsic worth’ (2002) Stell LR 299, 305 - 307.

5 Pace HLA Hart The concept of law (1960), Cornell writes: ‘Synchronisation points us to the real problem... How do we develop an institutional analysis which allows us not only to synchronise the competing rights of individuals, but also the conflicts between the individual and the community, and between different groups in society. The goal of a modern legal system is synchronisation and not coherence. Synchronisation recognises that there are competing rights situations and real conflicts between the individual and the community which may not yield a coherent whole. The conflicts may be mediated and synchronised but not eradicated.’ D Cornell ‘Pragmatism, recollective imagination, and transformative legal interpretation’ (1993) Transformations 23 35 - 36.


from the work of Kant) or communitarianism (drawn from the work of Rousseau or Marx), it provides a distinctly Southern African lens through which judges, advocates, attorneys and academics ought to determine the extension of the actual provisions of the basic law. It hardly seems controversial to ground the South African Constitution in the lived experience of South Africans so long as an ubuntu-based reading does no violence to the text.

That’s not an opinion. It’s the law.

When the Constitutional Court handed down Makwanyane in 1995, it clearly decided the matter in the spirit of ubuntu. The enabling environment at the Constitutional Court – of equals serving amongst equals with a deep sense of consideration and respect for each other’s views, and of a commitment to the ideal of a new constitutional democracy in which every contribution counts – allowed members of the Court to plant the seed of ubuntu jurisprudence. (The textual source of the seed can be found in the postscript of the interim Constitution). While most of the justices emphasised the violation of the right to human dignity or the right to freedom and security of the person as a basis for finding the death penalty infirm, several others based their findings upon accepted axioms of ubuntu:

Generally ubuntu translates as humanness. In its most fundamental sense it translates as personhood and morality. Metaphorically it expresses itself in umuntu, ngabantu ngabantu, describing the significance of group solidarity on survival issues so central to the survival of communities. While it envelops the key values of group solidarity, compassion, respect, human dignity, conformity to norms and collective unity, in its fundamental sense it denotes humanity and morality. Its spirit emphasises respect for human dignity making a shift from confrontation to conciliation.8

Members of the Makwanyane court found the death penalty repugnant because retribution and group catharsis as the bases for punishment are inconsistent with an ubuntu-based jurisprudence of reconciliation, restorative justice and democratic solidarity. As importantly, their findings should be understood as broadly representative of South African views regarding the moral underpinnings of the basic law. The presence of ubuntu as a guiding norm in the

8 S v Makwanyane 1995 3 SA 391 (CC), 1995 6 BCLR 665 (CC), 1995 2 SACR 1 (CC) para 308. In Makwanyane, Langa J writes: ‘Ubuntu captures, conceptually, a culture which places some emphasis on communality and on the interdependence of the members of a community. It recognises a person’s status as a human being, entitled to unconditional respect, dignity, value and acceptance from the members of the community such a person happens to be part of. It also entails the converse, however. The person has a corresponding duty to give the same respect, dignity, value and acceptance to each member of that community. More importantly, it regulates the exercise of rights by the emphasis it lays on sharing and co-responsibility and the mutual enjoyment of rights by all.’ (paras 224 - 225.)
interpretation of our basic law is essential for the legitimation of our legal system.\(^9\)

We ignore *ubuntu* at our own peril. Indeed, without saying much more on this subject, a growing sense of disjunction between the ideals of the Constitution and the lived experience of most South Africans warrants a reappraisal of the place of *ubuntu* in South African law. It is this *difference* between dignity – as espoused by Kant and other Western philosophers – and *ubuntu* as practiced by the majority of South Africans that animates Professor Cornell’s lead essay on the subject. We need to do more than infuse our dignity jurisprudence with a soupcon of *ubuntu*. The legitimation of the South African legal order depends upon our ability to synchronise these two closely related, but distinct terms. In *Khosa*, the Court offered the following South African gloss on the demands of dignity – one framed in a decidedly South African *lingua franca* – in finding that the State’s refusal to provide permanent residents with social welfare benefits constitutes a violation of the right to social security and the right to equality:

> Sharing responsibility for the problems and consequences of poverty equally as a community represents the extent to which wealthier members of the community view the minimal well-being of the poor as connected with their personal well-being and the well-being of the community as a whole. In other words, decisions about the allocation of public benefits represent the extent to which poor people are treated as equal members of society.\(^10\)

If dignity and *ubuntu* can be squared in such a fashion, and in the context of such a difficult case as the right to adequate access to social security, then one might ask, as Professor Cornell does – why has *ubuntu* been met by the academy and by the courts with such resistance?

We can identify two sources for this resistance: (1) the problem of translatability; and (2) the tension between radical reconstructions of *ubuntu*

---


\(^10\) See *Khosa v Minister of Social Development* 2004 6 SA 505 (CC), 2004 6 BCLR 569 (CC) (Mokgoro J) para 74. In *Khosa*, the Court goes beyond dignity as minimal respect and arrives at dignity as a collective concern. The Constitutional Court has discussed dignity as a collective responsibility in a number of its unfair discrimination decisions. See, eg, *Hoffmann v South African Airways* 2001 1 SA 1 (CC), 2000 11 BCLR 1211 (CC) para 43 (‘The interests of the community lie in the recognition of the inherent dignity of every human being and the elimination of all forms of discrimination.’) The Constitutional Court has written about dignity *qua* collective responsibility in the context of evictions and claims asserted under sec 26. See *Port Elizabeth Municipality v Various Occupiers* 2005 1 SA 217 (CC), 2004 12 BCLR 1268 (CC) para 18 (‘It is not only the dignity of the poor that is assailed when homeless people are driven from pillar to post in a desperate quest for a place where they and their families can rest their heads. Our society as a whole is demeaned when state action intensifies rather than mitigates their marginalisation.’)
Where dignity ends and ubuntu begins - answer to Drucilla Cornell

11 See, eg, T Metz ‘African ethics’ International encyclopedia of ethics (2010) (manuscript on file with authors). Metz maintains that ‘[t]here are kinds of communitarian and vitalist approaches to morality commonly held by sub-Saharan ethical theorists that international scholars should take seriously as genuine rivals to utilitarian, Kantian, contractarian, and care-oriented outlooks that dominate contemporary Euro-American discussion of right action ... Many friends of sub-Saharan morality would sum it up by saying what is most often translated (overly literally) as either “A person is a person through other persons” or “I am because we are”. One encounters such phrases in a variety of societies, ranging from those in South Africa to Kenya in East Africa and Ghana in West Africa. While these phrases do connote the empirical or even metaphysical idea that one needs others in order to exist, they also convey a normative outlook. In particular, personhood and selfhood, in much African moral thought, is value-laden, meaning that one’s basic aim as a moral agent should be to become a complete person or a real self. Or, using the influential term used among Zulu and Xhosa speakers in South Africa, one’s fundamental goal ought to be to obtain ubuntu, to develop humanness or live a genuinely human way of life. Insofar as a large swathe of sub-Saharan thought takes one’s proper ultimate end to be to become (roughly) a mensch, it may be construed as a self-realisation morality, not unlike Greek and more generally perfectionist standpoints ... However, unlike the self-realisation approaches that are dominant in the West, characteristic African versions are thoroughly relational or communitarian in the way they specify what constitutes one’s true or valuable nature. That is, most Western accounts of morality that direct an agent to develop valuable facets of her human nature conceive of there being non-derivative self-regarding aspects of it, such as properly organising one’s mental dispositions (Plato’s Republic) or understanding parts of the physical universe (Aristotle’s Nicomachean ethics). In contrast, a large majority of sub-Saharan conceptions of self-realisation account for it entirely in terms of positive relationships with other beings. In general, one develops one’s humanity just insofar as one enters into communal relationships with God and ancestors, viz, elderly and morally wise progenitors of a clan who are thought to have survived the death of their bodies and to continue to interact with those in this world (the ‘living-dead’ as they are often called) (citations omitted). Metz’s view of African ethics is not all sunshine and roses. See, eg, T Metz ‘African moral theory and public governance: Nepotism, preferential hiring and other partiality’ in MF Murove (ed) African ethics: An anthology for comparative and applied ethics (2009) 335.

its expositors could not convey with precision in the dominant language—English—and political culture—liberal capitalism—a term that was well understood in indigenous cultures throughout South Africa? We think that the failure of engagement reflects (in part) a general malaise on the part of academics and jurists. For some, this lassitude takes the form of a rather thin and potted post-structuralist or post-modernist account of other cultures—as in ‘it’s a black thing, we couldn’t possibly understand.’ But as Professor Cornell’s writing and her ubuntu project make clear, no good reason exists to countenance such relativism. Our point here is decidedly Davidsonian. What all relativists have in common—from Foucault or Derrida to Kuhn and Feyerabend—is a notion of ‘incommensurable conceptual schemes’. Davidson writes:

Philosophers of many persuasions are prone to talk of conceptual schemes. Conceptual schemes, we are told, are ways of organising experience: They are systems of categories that give form to the data of sensation; they are points of view from which individuals, cultures or periods survey the passing scene. There may be no translating from one scheme to another, in which case the beliefs, desires, hopes, and bits of knowledge that characterise one person have no true counterparts to the subscriber to another scheme. Reality is relative to a scheme: what counts as real in one system may not in another... Conceptual relativism is a heady and exotic doctrine, or would be if we could make sense of it. The trouble is, as so often in philosophy, that it is hard to improve intelligibility while retaining the excitement.

That’s about as charitable as Davidson can be in dismissing epistemological claims made by philosophers working within post-modern or post-structuralist frameworks. About the confusion associated with conceptual schemes, he argues:

Since knowledge of beliefs comes only with the ability to interpret words, the only possibility at the start is to assume general agreement on beliefs... The guiding policy is to do this as far as possible... The method is not designed to eliminate disagreement, nor can it; its purpose is to make meaningful disagreement possible, and this depends entirely on a foundation—some foundation—in agreement... We make maximum sense of the words and thoughts of others when we interpret them in a way that optimises agreement... Where does that leave the case for conceptual relativism? The answer is, I

13 A mainstay of the meta-theory of cultural anthropology and literary studies, relativism even managed to worm its way into the philosophy of science—P Feyerabend Against method (1975); P Feyerabend A farewell to reason (1987)—though its fair to say that most scientists do not take such relativism seriously.14 As D Davidson writes in his ‘Introduction’ Inquiries into truth and interpretation (1984) xiv (my emphasis): ‘So a theorem like “Schnee ist weiss” is true in the mouth of a German speaker if and only if “snow is white” has to be taken not merely as true, but as capable of supporting counterfactual claims. Indeed, given that the evidence for this law, if it is one, depends ultimately on certain causal relations between speakers and the world, one can say that it is no accident that “Schnee is weiss” is true if and only if snow is white; it is the whiteness of snow that makes “Schnee is weiss” true.’15 D Davidson ‘On the very idea of a conceptual scheme’ Inquiries into truth and interpretation (1984) 183 183.
think, that we must say much the same thing about differences in conceptual schemes as we say about differences in belief: we improve the clarity and bite of declarations of difference, whether of scheme or opinion, by enlarging the basis of shared (translatable) language or of shared opinion.16

The apposite conclusions to be drawn from Davidson’s analysis are (a) that no good reason exists to think that the large majority of beliefs held by members of another culture are somehow beyond our ken; and (b) that it is our intellectual job as academics and jurists, and our ethical responsibility as members of the same South African polity, to try to get our heads and hearts around words and concepts that undoubtedly feel (at first instance) foreign and uncomfortable.

The greater challenge, we believe, lies not in whether we can understand or can determine the exact contours of ubuntu, but in describing and defending a certain conception of ubuntu. Here we think that Professor Cornell would be well-served not only to defend ubuntu against its usurpation by the constitutional grundnorm of dignity, but to tackle head on conservative variants of ubuntu that would block her social democratic vision of a future South Africa. Bhe v Magistrate, Khayalitscha17 – in striking down traditional rules of male primogeniture – or Shilubana v Nwamitwa18 – in upholding changes to traditional norms that had previously permitted only male ascension to tribal authority – may seem like easy cases when analysed in terms of the existing jurisprudence of rights to equality of sex and gender and the right to dignity. But they are most assuredly not. Why?

Two reasons. First, a significant proportion of our country’s denizens view the Constitution ‘as read’ in cases such as Shilubana and Bhe as dramatically out of step with their ‘lived experiences’.19 Indeed, a growing contingent of scholars and members of social movements contend that our Western-style constitutional democracy is viewed as an imposition, as an extension of colonial rule and white power through law (as opposed to

16 Davidson (n 15 above) 197.
17 2005 1 SA 580 (CC), 2005 1 BCLR 1 (CC).
18 2009 2 SA 66 (CC), 2008 9 BCLR 914 (CC).
19 See T Madlingozi ‘The Constitutional Court, court watchers and the commons: A reply to Professor Michelman on constitutional dialogue, “interpretive charity” and the citizenry as sangomas’ (2008) 1 Constitutional Court Review 63. See also S Sibanda How constitutional democracy has blocked the promise of South Africa’s liberation movements (PhD, in progress 2011, University of the Witwatersrand); Khoisan Chief Jean Burgess ‘The Constitution does not speak to me or for me’ in The Constitution and the Masses Conference at the South African Institute for Advanced Constitutional, Public, Human Rights and International Law (Constitution Hill, 29 October 2010)(Chief Jean contends that not only does the Constitution’s failure to recognise Khoi and San languages as official languages exclude Khoisan people from participation in South Africa’s constitutional project, other laws that only permit persons who speak Khoisan languages to ‘speak’ for Khoisan communities further marginalise Khoisan peoples. As a result, she argues that such a flaw in the basic law functions as a barrier to Khoisan acceptance of the legitimacy of the Constitution.)
violent, fascist and expressly racist oppression.) Second, the process of ‘recollective imagination’ and ‘transformative legal interpretation’ of *ubuntu* into a set of beliefs and practices that will bring about a representative, or democratic socialist, or communitarian, South Africa polity cannot be assumed. The critique of *ubuntu* as a potentially conservative brake on this country’s emancipatory project cannot be cavalierly dismissed. If *ubuntu* is connected – however unreflectively – with practices such as male primogeniture, male ascension to leadership positions, male circumcision rites or compensation for teacher-student, male-on-female sexual violence, then such practices must be rooted out before *ubuntu* can lay claim to the mantle of revolutionary constitutional doctrine. We can agree with Professor Cornell’s desire to place *ubuntu* at the core of South African constitutionalism even as we insist that her efforts must convert those members of South African society who, when they speak of *ubuntu*, imagine a very different set of (conservative) political institutions and legal doctrines.

Professor Cornell’s reflections on *ubuntu* and dignity are neither inventions in the German Idealist tradition nor fantasies about the underlying philosophy of many South African communities. What Professional Cornell’s writings exhibit is a two-fold effort to excavate the profound and unique communitarian thought embedded in *ubuntu* and a rolling-up-of-the-sleeves exercise in determining the content of *ubuntu*-based forms of life throughout South Africa. Its also on exercise in speculative, philosophical anthropology. But it is more than that. By tying her unabashedly political project to classically anthropological work in the field, Professor Cornell promises the possibility of a fully worked out African ethics and politics. If there is a tension in Professor Cornell’s work, then it is a most natural tension between academic pursuits and practical politics. As a Kantian scholar of the first order, a long-standing student of *ubuntu* and a savvy analyst of politics in South Africa, the challenge before Professor Cornell is to get individuals and communities within this radically heterogeneous country singing off the same hymn sheet. No easy task.

As her paper reflects, Professor Cornell struggles mightily to close the gap between *ubuntu* and dignity – at the same time as she recognises the differences between both constitutional norms. Need we try so hard to reconcile the two? No one has suggested that we need to square *ubuntu* with equality or freedom, or reduce it entirely to community rights. We might do well to consider allowing these values to occupy their own separate spaces – closely aligned, but with different roles to play when we apply our minds to constitutional conflicts. Perhaps a Twainian twist is in order: *Ubuntu* and dignity do not map directly on to one another, but they do rhyme.

---

Given this added layer of complexity, we might ask whether Professor Cornell can make good on the promise of both her ubuntu and dignity projects? She has already more than started. Yet a sense of urgency in South African politics (yes, an incipient violence) lingers over her poignant manifestos. We can ask of the ubuntu project, as Hillel did of his brethren over two millennia ago, 'If not now, when?'
1 What does it mean to have a constitutional right?

The Criminal Law (Forensic Procedures) Bill 2 of 2009 was introduced into Parliament just prior to the 2009 election and was subsequently withdrawn without being considered by Parliament.¹ It was revived after the election and considered in detail by the Parliamentary Portfolio Committee on Police.² The Committee has since split the Bill into two parts. The first part deals with wide-ranging amendments to the law governing the obtaining of fingerprints, and it was adopted by the Committee in March 2010 and passed by the National Assembly in June.³ Consideration of part 2 of the Bill will take place later in the 2010 parliamentary session.

If it follows the form of the Bill originally introduced in 2009,⁴ then part 2 of the Forensic Procedures Bill will amend the South African Police Service Act 68 of 1995 and permit the establishment of a national DNA database. This will consist of DNA profiles (popularly known as a ‘DNA fingerprint’) derived from DNA samples that have been collected from crime scenes, from arrested persons, from convicted offenders and from volunteers. In addition to DNA profiles, the database will also hold, ‘where applicable and scientifically possible’, the physical DNA samples for possible use if ‘significant technological advances’ have been made since the compilation of the DNA profile to which they relate.

³ The most important change proposed by this part of the Bill is to allow the police to perform comparative searches of other, non-forensic, fingerprint databases such as the databases of the Departments of Home Affairs and of Transport. See clause 6 of the Bill.
⁴ Some changes can be expected. The reasons for the Committee delaying consideration of the DNA-database aspect of the original Bill were concerns about the readiness of the police to implement it and concerns about the constitutionality of some of its proposed measures. See ‘DNA database delayed’ IT Web 5 July 2009.
If enacted in this form, this legislation is likely to be challenged as an infringement of the right to privacy. The right to privacy in the South African Constitution includes protection of ‘informational privacy’, an individual’s interest in restricting the collection, use and retention of information about themselves.5

Such a challenge has been presaged by the case of *S and Marper v United Kingdom*, in which the European Court of Human Rights considered the conformity of a similar DNA database in the United Kingdom to the European Convention on Human Rights.6 The Court found that mere storing of data relating to the private life of an individual amounted to an interference with the European Convention’s right to respect for private life, the interference being particularly serious when that data was an individual’s cellular sample and DNA profile.7

That, of course, was not the end of the story in the European decision, just as it would not end the argument if a similar challenge were to be brought under the South African Constitution. In Europe, the right to privacy is subject to the limitation that it may be interfered with by a public authority in accordance with the law and is necessary in a democratic society ... for the prevention of disorder or crime ...’.8 Strong arguments can be made for the necessity of the interference, the flavour of which is given in this extract from the court’s summary of the UK government’s case:

The interference was necessary and proportionate for the legitimate purpose of the prevention of disorder or crime and/or the protection of the rights and freedoms of others. It was of vital importance that law enforcement agencies took full advantage of available techniques of modern technology and forensic science in the prevention, investigation and detection of crime for the interests of society generally. They submitted that the retained material was of inestimable value in the fight against crime and terrorism and the detection of the guilty and provided statistics in support of this view. They emphasised that the benefits to the criminal-justice system were enormous, not only permitting the detection of the guilty but also eliminating the innocent from inquiries and correcting and preventing miscarriages of justice.9

Though these arguments did not persuade the European Court, they had considerably more success in the United Kingdom courts, persuading the Administrative Court and majorities of the Court of Appeal and the House of Lords that the interference with privacy was justifiable. The judgment of Waller LJ sums up the approach of the UK courts: The interferences with privacy by the retention of profiles and samples ‘are not great, and such as they are they are outweighed by the benefits in achieving the aim of

---

5 *Mistry v Interim National Medical and Dental Council of South Africa* 1998 4 SA 1127 (CC).
7 Marper (n 6 above) 67.
8 Art 8(2) of the European Convention on Human Rights.
9 Marper (n 6 above) 91.
prosecuting and preventing crime’.10 By contrast, the unanimous conclusion of the European Court was that the balance went the other way. It was found that ‘the blanket and indiscriminate nature of the powers of retention of the fingerprints, cellular samples and DNA profiles of persons suspected but not convicted of offences, as applied in the case of the present applicants, fails to strike a fair balance between the competing public and private interests’.11

Under the South African Constitution, it is certainly an arguable proposition that retention of DNA data and cellular samples in a forensic database is constitutionally permissible, despite its infringement of the right to privacy. The argument in a constitutional challenge is likely to proceed in much the same way as outlined above. What then does it mean to say, as section 14 of the Constitution does, that ‘everyone has the right to privacy’? What exactly is it that this declaration achieves if the right can simply be balanced against the interests of crime detection and prevention and, as the decisions of the UK courts illustrate, may be outweighed by these policy goals? If rights can be weighed against competing public interests justifying their restriction, and if they are sometimes outweighed by those interests, then it is not clear what is gained by having a right. This conception of rights is at odds with a conception of rights as having special strength and significance, with the idea that, if they are to be taken seriously, rights should not simply be traded against competing policy goals.12

2 The ubiquity of balancing

The principle of proportionality which is the hallmark of the jurisprudence of the European Court of Human Rights and of the UK courts under the Human Rights Act is also the principle governing the practice of the South African courts under the 1996 Constitution. In the latter case, this is because of the presence of the limitation section in the South African Constitution, which provides that the rights in the Bill of Rights can be limited by law if there are good reasons for doing so.13 The limitation section sets out criteria for determining what constitutes a good reason for limiting rights. These are the criteria in the part of the general formula beginning with the words ‘the limitation is reasonable and justifiable in an open and democratic society based on human dignity, equality and freedom’ and the first three limitation

10 R (Marper) v Chief Constable of the South Yorkshire Police [2002] EWCA Civ 1275, quoted by the ECtHR (n 6 above) 13.
11 Marper (ECtHR) (n 6 above) 125.
12 M Kumm ‘Political liberalism and the structure of rights’ in G Pavlakos (ed) Law, rights and discourse: The legal philosophy of Robert Alexy (2007) 131. According to Kumm, the general conception of rights in the liberal tradition is that rights may not be overridden by ordinary considerations of policy. This formulation unites Dworkin’s conception of rights as ‘trumps’, Rawls’s notion of the priority of the right over the good and Habermas’s conception of rights as having a deontological character that withdraws them from participation in a cost-benefit analysis.
Chapter 11

factors – inviting consideration of the limited right, the extent of its limitation and the purpose of the limitation. The section also requires consideration of whether the limitation will achieve the purpose it is designed to achieve and whether there are other ways in which the purpose can be achieved without restricting rights, or by restricting them less severely. Because it is a general limitation section, all the rights are limitable and all are limitable according to the same set of justifying criteria.

The structure of rights analysis in a Bill of Rights with a general limitation section, the Constitutional Court held, in its first encounter with the current section’s predecessor in *S v Makwanyane*, ‘involves the weighing up of competing values, and ultimately an assessment based on proportionality’. ‘Proportionality’, the Court continued, ‘calls for the balancing of different interests’.14 It has been saying more or less the same thing ever since. To use a more recent example, *Brümmer v Minister for Social Development*,15 the court gave the following summary in a nutshell of the limitation enquiry:

In assessing whether the limitation ... is reasonable and justifiable under section 36(1), regard must be had to, among other factors, the nature of the right limited; the purpose of the limitation, including its importance; the nature and extent of the limitation; the efficacy of the limitation, that is, the relationship between the limitation and its purpose; and whether the purpose of the limitation could reasonably be achieved through other means that are less restrictive of the right in question. Each of these factors must be weighed up but ultimately the exercise is one of proportionality which involves the assessment of competing interests.16

There are dozens of similar summaries in the case law. That the limitation enquiry entails, in part, the ‘weighing up’ or ‘balancing’ of competing considerations can safely be said to be a trite proposition, an article of faith in South African constitutional law. To be more precise: what is trite and, as I will elaborate, controversial about this understanding of the limitation enquiry, is that it countenances a head-to-head conflict between rights and competing public interests and the possibility that rights will sometimes be outweighed by those interests. The other part of the limitation analysis, entailing an enquiry into the relationship between the limitation and the purpose for it – between means and ends – is, for current purposes at least, a relatively uncontroversial aspect of the overall proportionality enquiry. To take the *Brümmer* decision as a conveniently simple example of this distinction, the Court’s limitation analysis proceeds as follows:

14 *S v Makwanyane* 1995 3 SA 391 CC 104 (Chaskalson P).
15 2009 6 SA 323 CC.
16 *Brümmer* (n 15 above) 59 (Ngcobo J).
(a) A 30-day time bar on appeals to court against a refusal of a request for access to information is a limitation of the right of access to court and the right of access to information.\textsuperscript{17}

(b) The right of access to court is important; it is ‘foundational to the stability of an orderly society’.\textsuperscript{18}

(c) The right of access to information is also important; the ‘importance of this right too, in a country which is founded on values of accountability, responsiveness and openness, cannot be gainsaid’.\textsuperscript{19} The right is fundamental to the realisation of other rights such as freedom of expression.\textsuperscript{20}

(d) Time-bar provisions are important, avoiding delays in litigation that hamper the interests of justice.\textsuperscript{21} They give effect to practical considerations that arise with the passage of time between the occurrence of the cause of action and the litigation: lost evidence, disappeared or forgetful witnesses. They seek to reduce the costs and burdens associated with responding to litigation long after the fact that give rise to it.

(e) Accordingly, in previous decisions, the Court has upheld time bars as constitutional, but only if they nevertheless provide a litigant an adequate and fair opportunity to bring an application to court.\textsuperscript{22}

(f) In access to information litigation against the state the usual considerations motivating time bars are not present or are only minimally present. The litigation is about a record in the possession or under the control of the state and the evidence of witnesses is seldom required. There is little burden in preserving a contested record until the litigation is completed.\textsuperscript{23}

(g) The experience of litigants and reference to comparable legislation suggest that the 30-day limit in the Promotion of Access to Information Act is too short to provide an adequate and fair opportunity to bring an appeal against refusal of access to court. A six-month limit would serve the interests of the state but would not unduly restrict access to court by an aggrieved requester.

In the scheme above it is only holdings (b), (c) and (d) that constitute a balancing enquiry properly so-called, an enquiry into whether the purposes served by time bars can justify the limitation of the rights of access to court and access to information. The operative part of this particular decision is however holdings (e) and (g), which result from an enquiry into means and ends. In the end, the time bar is found to be disproportionate in the sense that a thirty-day limit does more harm than it needs to. A six-month limit on the other hand would achieve the beneficial purposes of a time bar with a comparably less severe limitation of the rights.

\textsuperscript{17} Brümmer (n 15 above) 57. The time bar is imposed by sec 78(2) of the Promotion of Access to Information Act 2 of 2000.

\textsuperscript{18} Brümmer (n 15 above) 61 quoting Chief Lesapo v North West Agricultural Bank 2000 1 SA 409 (CC) 22.

\textsuperscript{19} Brümmer (n 15 above) 62.

\textsuperscript{20} Brümmer (n 15 above) 63.

\textsuperscript{21} Brümmer (n 15 above) 64.

\textsuperscript{22} Brümmer (n 15 above) 52.

\textsuperscript{23} Brümmer (n 15 above) 65 - 66.
3 Criticism of balancing

Is the unswerving commitment of the South African courts to balancing appropriate, given that, according to its many critics, balancing has serious limitations? In their chapter on ‘Limitations’ in the second edition of Constitutional law of South Africa, Woolman and Botha criticise not only the particular form of balancing enquiry undertaken in South African limitations analysis, but also of the foundational idea that rights can and should be balanced against conflicting considerations. Woolman and Botha have distilled a voluminous critical literature into a useful compendium of the arguments against the use of balancing as a metaphor to describe the process of conflict resolution in limitations analysis. According to the authors, balancing has four principal defects and the continued deployment in limitation analysis of this ‘bad metaphor’ leads limitation jurisprudence seriously astray. The defects are the following:

(a) Incommensurability. The metaphor of ‘balancing’ and its partner ‘weighing-up’ suggest that rights and the public interest in their limitation are commensurable, measurable by the same metric (or, to keep to the balancing metaphor, scale). Some sort of common metric is necessary to be able to say that one value is better than the other, or that they are of equal value. But, Woolman and Botha argue that, more often than not, there is no such metric: ‘Human goods are often incommensurable.’

(b) Subjectivity and arbitrariness. The absence of an objective, external metric for the comparison and ordering of competing values creates the danger of subjectivism, that judges will use their own personal metric when balancing. Unchecked judicial discretion of this sort is arbitrary, encouraging adjudication either ungoverned by rules or taking place ‘unreflectively, according to customary standards and hierarchies’.

(c) Incrementalism and conservatism. Balancing, Woolman and Botha argue, lends itself to a ‘cautious, incrementalist approach to constitutionality inspired judicial law-making. The balancer is inclined to restrict her finding to the case at hand, as the next case may, ostensibly, require that a different balance be struck.’

25 Woolman & Botha (n 24 above) 34 - 94.
26 See J Raz ‘Value incommensurability: Some preliminaries’ (1985 - 1986) 86 Proceedings of the Aristotelian Society, New Series 117, 117: ‘A and B are incommensurate if it is neither true that one is better than the other nor true that they are of equal value.’
27 Woolman & Botha (n 24 above) 34 - 96.
28 Woolman & Botha (n 24 above) 34-99 - 34-100.
31 Woolman & Botha (n 24 above) 34-100 - 34-101. The objections to a ‘one case at a time’ approach to constitutional adjudication are more fully set out in Woolman ‘The amazing, vanishing bill of rights’ (2007) 124 SALJ 762.
Balancing and the limitation of rights in the South African Constitution 257

(d) Quasi-scientificity. The balancing metaphor and related language suggests that the process of limitation analysis is technical, even mechanical. The result, as Aleinikoff puts it, ‘is simply read off the machine. Scientific balancing decisions are neither opinions nor arguments that can engage us; they are demonstrations’. This might comfort critics troubled by the spectre of excessive judicial discretion, but it is not an accurate description of the actual practice and also ‘avoids dialogue about important moral and political issues’. A fifth criticism of balancing, not included in Woolman and Botha’s compendium, can be added to this list. This is the criticism, rehearsed in the opening section of this article, that balancing fails to take rights seriously, that it fails to accord to rights a necessary special priority. We can take Habermas as a representative of this line of thought.

(e) Taking rights seriously. Habermas’s critique of balancing is a response to Robert Alexy’s defence of the practice in his Theorie der Grundrechte (1986). In the postscript to the English translation of this work, Alexy summarises Habermas’s argument as follows:

… ‘goal-oriented weighting’ leads to the possibility that ‘individual rights can be sacrificed at times to collective goals’. But then constitutional rights would lose their firmness, which can only be guaranteed by way of a strict deontological structure, that is by having the character of rules. By contrast, giving them the character of principles destroys a ‘fire wall’. ‘For if in case of collision all reasons can assume the character of policy arguments, then the fire erected in legal discourse by a deontological understanding of legal norms and principles collapses’.

Powerful though the case against balancing seems to be, this has not discouraged the South African courts from continuing to practise it. Nor is the practice confined to this jurisdiction; as a component of proportionality, balancing could be considered ubiquitous in contemporary constitutional law.

32 Aleinikoff (n 29 above) 993, quoted in Woolman & Botha (n 24 above) 34-101.
33 Woolman & Botha (n 24 above) 34-101.
36 As Woolman observes in ‘Vanishing Bill of Rights’ (n 31 above) 783 n 41: ‘I have offered several related critiques of the court’s reliance on the metaphor of “balancing” when undertaking limitation analysis. The court’s subsequent jurisprudence reflects neither an engagement with those critiques nor any effort to offer a coherent account of what it often calls “proportionality” analysis’.
4 The case for balancing

Missing from Botha and Woolman’s account is the powerful counter-case in favour of balancing. The case is made in Alexy’s descriptive work on the practice of the German Constitutional Court, a practice that has been characterised since the inception of the court by a commitment to the principle of proportionality and to its sub-principle, balancing. Though Alexy is cautious about claiming the accuracy of his theoretical conclusions about the nature of rights adjudication only in relation to the jurisprudence of the German Court, his account of rights obviously has more general application to any jurisdiction to which proportionality and balancing has spread. What, then, is Alexy’s answer to the objections to balancing that have been set out above?

We need to begin with an overview of Alexy’s argument about the nature of constitutional rights, as the principle of proportionality and balancing are necessary attributes of this particular conception of rights. One can begin with Dworkin’s famous debate with Hart about positivism’s ‘model of rules’. Law, Dworkin showed, consisted of more than just a set of valid legal rules but contained also ‘standards that do not function as rules, but operate differently as principles, policies and other sorts of standards’. Dworkin illustrated the existence and operation of one of these standards – principles – by reference to the 1889 case of *Riggs v Palmer*, which posed the question as to whether an heir named in the will of his grandfather could inherit when he had murdered his grandfather. The decision entailed the clash of competing legal principles – namely, that no one should profit from his wrong versus the principle that wills should be upheld. The result in *Riggs* was in favour of the first principle, the murderer did not inherit. But this decision did not invalidate the second principle, which remained good law for use in a future case. This showed, Dworkin reasoned, that principles were structurally different to rules, lacking the ‘all or nothing’ quality of the latter and having, instead, a ‘dimension of weight’ which permitted conflicting principles to be maintained in a legal system without cancelling each other out:

> When principles intersect ... one who must resolve the conflict has to take into account the relative weight of each. This cannot be, of course, an exact measurement, and the judgment that a particular principle or policy is more important than another will often be a controversial one. Nevertheless, it is an integral part of the concept of a principle that it has this dimension, that it makes sense to ask how important or how weighty it is.

Alexy’s work builds from this relatively modest foundation a general classification of constitutional rights norms into rules and principles: ‘the
distinction between rules and principles is a basic pillar in the edifice of constitutional rights theory'. His next step is the characterisation of principles as 'optimisation requirements', 'norms which require that something be realised to the greatest extent possible given the legal and factual possibilities'. Unlike rules, which are definitive commands, norms 'which are always either fulfilled or not', principles can be 'satisfied to varying degrees' depending on what is factually and legally possible in a particular case. A constitutional rule is therefore something like section 49(1) of the Constitution – ‘The National Assembly is elected for a term of five years’ – or section 79(5) – ‘If the Constitutional Court decides that the Bill is constitutional, the President must assent to it and sign it’ – or section 5 – ‘The national flag of the Republic is black, gold, green, white, red and blue ...’ The application of rules like these entails subsumption. If a rule is valid and applicable, it is required to do exactly what it demands be done. If this is done, the rule is complied with; if this is not done, the rule is not complied with. A constitutional principle, on the other hand, is relatively easily identified since the Constitution self-consciously characterises some of its provisions as such: for example, the 'Principles of Co-operative Governance' listed in section 41, or the 'Basic Values and Principles Governing Public Administration' in section 195. It is also exemplified by any of the rights in the Bill of Rights, since each right is notionally subject to the possibility of limitation in terms of the general limitation section. ‘Everyone has the right to privacy’ seems on the face of it to state a rule, but in fact it does not since we know that it does not state a categorical prohibition on restrictions of privacy; privacy can in fact be restricted in the service of countervailing, constitutionally-recognised considerations such as public safety or the detection and prosecution of crime or to give effect to competing rights such as freedom of expression. Constitutional rights therefore have the structure of principles since they state what ought to be realised to the greatest extent relative to what is legally and factually possible. The legal possibilities are determined largely by competing principles, the factual by the facts of the case to be decided.

Conflict between rules can arise and is resolved in one of two ways: Either an exception is read into one of the rules, or one of the rules is declared invalid.

Competing principles are not resolved this way but instead by operation of a characteristic of principles that, as Dworkin noticed, is not possessed by rules – the dimension of weight:

42 R Alexy Theory of constitutional rights (n 35 above) 44. There are no other norms, according to Alexy, ‘Every norm is either a rule or a principle’ 48.

43 Alexy Theory of constitutional rights (n 35 above) 48 - 49.

44 Subsumption involves the comparison of abstract norms with the concrete facts in a given case to determine the applicability of the norms. R Alexy ‘On balancing and subsumption: A structural comparison’ (2003) 16 Ratio Juris 433.

45 ‘Notionally’ because the general limitation section is not quite as general as it seems. See Woolman & Botha (n 24 above) 34.4 on the problem of internal modifiers in the formulation of some of the rights. There are also, here and there, rules in the Bill of Rights, eg sec 25(9) ‘Parliament must enact the legislation referred to in subsection (6)’.

46 Alexy Theory of constitutional rights (n 35 above) 49.
If two principles compete, for example if one principle prohibits something and another permits it, then one of the principles must be outweighed. This means neither that the outweighed principle is invalid nor that it has to have an exception built into it. On the contrary, the outweighed principle may itself outweigh the other principle in certain circumstances. In other circumstances the question of precedence may have to be reversed. This is what is meant when it is said that principles have different weights in different cases and that the more important principle on the facts of the case takes precedence. Conflicts of rules are played out at the level of validity; since only valid principles can compete, competitions between principles are played out in the dimension of weight instead.47

The metaphor of weight suggests that the process of determining which principle outweighs the other in a particular case necessarily entails balancing. This turns out to be so: For Alexy, just as subsumption is the form of application of rules, so balancing is the specific form of the application of principles. One can take again as an example the decision in Brümmer on the constitutionality of time bars. Simplifying the holding to emphasise the conflict of principles one could formulate the contention as follows:

Principle 1: Everyone has the right of access to court. (P1)

Principle 2: The administration of justice requires the prevention of ‘inordinate delays in litigation’.48 (P2)

The conflict is resolved, according to Alexy, by establishing a conditional relation of precedence between the two principles: ‘On any set of concrete facts principle P1 has greater weight than the opposing principle P2 when there are sufficient reasons for supposing that in the circumstances of the concrete case, P1 takes precedence over P2.’49 As we have seen, the Constitutional Court’s analysis proceeds along just these lines, holding that P1 wins because of the following conditions: The legislation imposes an extremely restrictive time bar making litigation all but impossible without much corresponding benefit to the administration of justice, given the typical form of and typical evidence at stake in access to information litigation. The result can be stated as a rule, a rule that the constitutional right of access to court is breached under the conditions stated. This is the Law of Competing Principles: ‘The circumstances under which one principle takes precedence over another constitute the conditions of a rule which has the same legal consequence as the principle taking precedence.’50

One further element of Alexy’s account of rights should be covered here: his identification of a constitutive rule for balancing exercises called the ‘Law of Balancing’. One can formulate the reasoning of the court in Brümmer

47 Alexy Theory of constitutional rights (n 35 above) 50.
48 Brümmer (n 15 above) 64.
49 Alexy Theory of constitutional rights (n 35 above) 52 - 53.
50 Alexy Theory of constitutional rights (n 35 above) 54.
in a way that clearly brings out the dimension of weight involved in the balancing analysis: a short time-bar requires stronger reasons of practicality and the avoidance of inconvenience to justify it than a relatively generous time-bar does. So, a thirty-day limit could not be justified by the relatively minor benefits it accorded to defendants in access to information cases, while a six-month limit could. This reasoning follows the model of the Law of Balancing, which goes like this: ‘The greater the degree of non-satisfaction of, or detriment to, one principle, the greater must be the importance of satisfying the other.’\(^{51}\) This statement shows that balancing can be broken down into three stages. The first involves establishing the degree of non-satisfaction of, or detriment to, Principle 1. The second stage involves assessing the importance of satisfying the competing principle. The third involves establishing whether the importance of satisfying the competing principle justifies the detriment to, or non-satisfaction of the first.\(^{52}\)

5 Objections to the principle theory and to balancing

According to Alexy, the resolution of conflict between principles requires balancing. Balancing is therefore an unavoidable practice of constitutional adjudication. If so, it is a practice that is doomed to incoherence unless the objections to balancing that were outlined above can be answered. What follows is a preliminary attempt to meet the case against balancing informed by Alexy’s conception of balancing that has been outlined above. I will however, for present purposes, gloss over two of the criticisms made by Woolman and Botha. One of these is the charge that balancing encourages conservatism, a charge made rather faintly by the authors who admit that balancing has also been associated with progressive courts and decisions. As to the charge that balancing can be associated with a predilection for one-case-at-a-time incrementalism, and though this is probably true, it does not seem to me to be a bad thing.\(^{53}\)

Incommensurability. As we have seen, the results of balancing entail, for purposes of making a decision in a concrete case, a choice to prioritise one principle over another or even over multiple competing principles: This is the Law of Competing Principles described above. To be able to make such a choice the competing principles must be comparable, capable of being weighed on either side of the same scale. But, and this is the spectre raised by Woolman and Botha, what if they are not so capable and are in fact not commensurable? If so, the choice of one over the other is unjustifiable and should not have been made. But recall Woolman and

\(^{51}\) Alexy Theory of constitutional rights (n 35 above) 102.
\(^{52}\) Alexy Theory of constitutional rights (n 35 above) 401.
Botha’s contention that ‘human goods are often incommensurable’. How much work is being done by the weasel word – ‘often’? We might ask how often are the principles that conflict with each other in constitutional cases in fact incommensurable? Take Brümmer as an example, with its conflict between, on the one hand, a right of access to court, valued because it promotes the peaceful resolution of disputes and avoids resort to self-help and, on the other, a principle promoting order and efficiency in the administration of justice. Or take Marper as another example, which pits the right to privacy, valued, amongst other things, for its dignity-reinforcing effects, against the public interest in the efficient and error-free detection of crime. Deciding which principle to prefer on the facts of the case might be a more or less difficult decision, but it does not seem to require the weighing of incompatible options like, to borrow an example from the incommensurability literature, the decision whether to become a lawyer or a clarinettist. Decision making about these instances of the competition of principles is made on the basis of a common point of view: that is, the point of view of the Constitution with its range of values all requiring simultaneous accommodation to the greatest degree possible. I will say more about this in the treatment of the weight formula below.

Subjectivity and arbitrariness. Confronting the necessity to decide between arguments in favour of rights and in favour of their limitation requires, according to Woolman and Botha, ‘constitutional interpretation – and extended and reflected engagement with the meaning of the constitutional text’. The danger with balancing is that it encourages such decisions to be made instead on the basis of the ‘subjective preferences of individual judges’; the contention is nicely exemplified by the Marper case with its stark division of opinion between the English courts and the European Court on the proper balance between privacy and the claims of the criminal justice system. Along similar lines, Habermas has argued that balancing is ‘arbitrary’ and lacking in ‘rational standards’. Because there are no rational standards for balancing, ‘weighing takes place either arbitrarily or unreflectively, according to customary standards and hierarchies’. This is a serious charge. ‘If balancing or weighing were incompatible with correctness, objectivity, and justification’, Alexy admits, ‘it would have no place in constitutional law’. Recall the three stages of the balancing enquiry outlined above. The irrationality criticism would be justified if, as Alexy puts it ‘... it were not possible to make rational judgments about, first,

54 Woolman & Botha (n 24 above) 34 - 96.
55 See J Raz ‘Value incommensurability’ (n 26 above) 126.
56 Woolman & Botha (n 24 above) 34 - 100.
57 Habermas Between facts and norms (n 30 above) 259. ‘Customary standards’ means, at least in Alexy’s interpretation, something like the doctrine of precedent. See Alexy Theory of constitutional rights (n 35 above) 405.
intensity of interference, secondly, degrees of importance, and thirdly their relationship to each other. As principles, constitutional rights would then ... permit any solution.\textsuperscript{59}

Instead, according to Alexy, balancing as a procedure generates a predictable outcome 'not in every case, but in at least some cases, and that the class of these cases is interesting enough to justify balancing as a method'.\textsuperscript{60}

Take, again, \textit{Brümmer} as an example of a case in which a rational judgment about intensity of interference and degree of importance are able to produce a rationally defensible outcome. A three-year time bar on litigation seems, in obvious ways, a relatively minor interference with the right of access to court, with a one-month bar being at the other end of the scale of intensity and a six-month bar somewhere in the middle. Similarly, a complete prohibition on litigating (an ouster) would be a more severe infringement than the most parsimonious of time bars. A rough scale of interferences – light, moderate, severe – can be constructed and rational assignations can be made on this scale.\textsuperscript{61} As to the importance of the purpose of the limitation, the interests of justice in protecting the civil legal system from the inconvenience and uncertainty caused by delays in bringing litigation can plausibly be said to have some weight as a principle, but to be not particularly weighty. A case, such as \textit{Brümmer}, where the reasons for a moderate infringement of P1 is the relatively unweighty P2, the outcome is predictable and therefore not arbitrary. It is also not unreflective, in the sense that it is the outcome of reasoned argument, argument that, though it relies on precedent, does not consider precedent to be simply dispositive of the case without the need for further justification.

\textit{Brümmer} is a relatively simple case. Harder cases may involve greater possibility of variation in the conclusions drawn at the various stages of balancing, as is illustrated by the difference of opinion between the UK courts (and among some of the judges on those courts) and the European Court on the proper balance between privacy and an effective criminal justice system. This variability, this open-endedness, does not, however, mean that the balancing process is irrational: 'It is not as if we normally think of other open-ended decision-making approaches as irrational just because they involve some likelihood of variability.'\textsuperscript{62}

Alexy shows that one can go further and reduce the balancing enquiry to a formula which, though it allows for the exercise of judgment on the

\textsuperscript{59} Alexy \textit{Theory of constitutional rights} (n 35 above) 401.
\textsuperscript{60} Alexy \textit{Theory of constitutional rights} (n 35 above) 402.
\textsuperscript{61} Alexy \textit{Theory of constitutional rights} (n 35 above) 402.
values to be given to the variables of the formula, generates a predictable outcome in concrete cases. This is the weight formula, a simplified version of which looks like this:

\[
W_{ij} = \frac{I_i}{I_j}
\]

As Alexy explains the formula

\(I_i\) stands for the intensity of interference with the principle \(P_i\) ... \(I_j\) stands for the importance of satisfying the competing principle \(P_j\)... Finally, \(W_{ij}\) stands for the concrete weight of the principle whose violation is being examined – in our case, that of \(P_i\). The Weight Formula gives expression to the point that the concrete weight of a principle is a relative weight. It does this, in the simplest form, by defining the concrete weight as the quotient of the intensity of interference with this principle \((P_i)\) and the concrete importance of the competing principle \((P_j)\).

The operation of the formula can be illustrated by assigning numbers to the variables corresponding to the scale of intensity outlined above – light, moderate, severe. A geometric scale is used – \(2^0, 2^1, 2^2\), that is, \(1, 2, \) and \(4\). The priority of \(P_i\) is expressed by a concrete weight greater than \(1\); the priority of \(P_j\) by a concrete weight less than \(1\).

To use Brümmer as an example, the intensity of the infringement \((I_i)\) with the right of access to court \((P_i)\) is serious \((s)\); the importance of the interests of justice in the avoidance of inconvenience \((P_j)\) of the defendant \((I_j)\) – given the nature of access to information litigation is light \((l)\). Inserting the corresponding values of the geometric sequence for a severe infringement and a light importance then the concrete weight of \((W_{ij})\) is, in this case, \(4/1\), that is, \(4\), indicating the clear priority of \(P_i\) in this case. Even if greater weight is given to the purposes of the infringement, calling it moderately important would give a result of \(2\), which would still not be sufficient to outweigh the infringement of the right.

---


64 Alexy explains the choice of this scale in an instructive footnote: ‘The greatest advantage of the geometric sequence consists in its providing for the best representation of the over-proportional increase of the power of rights as correlated with an increasing intensity of interference, a fact that serves as the basis for the refutation of the objection concerning the dissolution of the power of constitutional rights’. Alexy ‘The construction of constitutional rights’ (n 63 above) n 36.
Taking rights seriously. Giving rights the character of principles and the practice of balancing that this conception of rights implies, so runs Habermas’s critique, softens them, destroying their essential character as rules: ‘The fire wall erected in legal discourse by a deontological understanding of legal norms and principles collapses.’ This understanding of rights stems from the Kantian conception of rights as having priority over the pursuit of collective goals. But treating rights as principles gives them no special priority, instead rights and policy goals ‘compete on the same plane within the context of proportionality analysis’.

Alexy’s response is rehearsed in the discussion of the weight formula above. The balancing enquiry does not give rights priority over competing principles, but is capable of giving them additional weight. This, coupled with procedural advantages for the applicant in constitutional cases, presumptions, burdens of proof and the two-stage methodology of rights analysis gives considerable advantages to the rights claimant. Proportionality and balancing, properly practiced, are not an all-things-considered analysis.

6 Conclusion

Balancing is a metaphor used to describe the process of legal reasoning in rights cases. It is an apt metaphor to describe a process that entails finding a resolution to a problem of competing principles by the ascription of relative weights to the competitors. This is a process made inevitable by the particular structure of the South African Bill of Rights; a generous catalogue of rights, each of which is expressly qualified by the possibility of limitation in the service of countervailing considerations. Balancing ought then to be defended and the nature of the process, its ‘internal justification’ as Alexy terms his deductive scheme, should be made visible in the interests of laying bare the legal reasoning behind decisions in constitutional rights cases.

65 There is more to say about this issue, but I will leave the heavy lifting to the article by Bilchitz in this volume. See Bilchitz ‘Does balancing adequately capture the nature of rights’ in S Woolman & D Bilchitz (eds) Is this seat taken? Conversations at the Bar, the bench and the academy about the South African Constitution (2012) 267.

66 M Kumm ‘Political liberalism and the structure of rights’ (n 12 above) 142.

67 There is, accordingly, much force in Woolman & Botha’s (n 24 above) criticism of the particular form of balancing and proportionality enquiry engaged in by the Constitutional Court which inclines on occasion to such an analysis.
Fifteen years into our constitutional democracy, it is accepted as a trite doctrine of constitutional law that fundamental rights are capable of limitation. Where the constitutionality of legislative or executive conduct is impugned, the Bill of Rights is commonly regarded as requiring a two-stage inquiry: First, has a right been violated in these circumstances? And secondly, if there has been a violation, can it be justified in terms of section 36(1) of the Constitution, the general limitation clause?¹

Whilst this is now a standard feature of South African constitutional law, the limitation of rights and the way it is conceived is of great significance and affects the very meaning of what it is to have a right in the South African Constitution.

In his paper that defends a balancing approach to the limitation of rights, Currie gives the example of *Marper v United Kingdom* which dealt with whether the police’s keeping of a database of DNA samples of individuals violated fundamental rights in the European Convention of Human Rights and, in particular, the right to privacy contained therein.² The same issue, he points out, could soon be before South African courts as parliament is considering a Bill dealing with the same issue.³ Given current South African law on the right to privacy, it is likely that such legislation would be held to violate this right. Nevertheless, the matter does not end there and there is a strong likelihood, particularly given the high crime rate in South Africa, that courts might hold

---

¹ This two-stage approach was outlined, eg, in *S v Makwanyane* 1995 3 SA 391 (CC) para 100. See also S Woolman & H Botha ‘Limitations’ in Woolman et al Constitutional law of South Africa (2006) (2nd ed) 34-3 - 34-4.
² *S and Marper v United Kingdom* (Application nos 30562/04 and 30566/04) (ECtHR 2008-12-04).

* Associate Professor, University of Johannesburg and Director, South African Institute for Advanced Constitutional, Public Human Rights and International Law (SAIFAC).
that such a violation is justified for purposes of solving crimes and holding perpetrators to account. If this is so, then it is unclear, as Currie puts it, what the right to privacy in fact achieves where it ‘can simply be balanced against the interests of crime detection and prevention and … outweighed by these policy goals?’

In his paper, Currie goes on to outline the way in which the South African Constitutional Court has dealt with the limitation of rights in terms of proportionality and a balancing of interests. He then raises various criticisms that have been advanced against the notion of balancing by Botha and Woolman.

Surprisingly, these authors have not been challenged until this point. Currie seeks to use the impressive work of Alexy to mount a defence of balancing, seeking to demonstrate the way in which Alexy’s theory may help answer some of the challenges posed by Botha and Woolman.

Alexy’s theory is rich and has great explanatory power: It is indeed a worthwhile project for academics to engage with his oeuvre and draw out the possibilities that his view holds for South African constitutional law. In this article, I shall consider two objections to Alexy’s theory. The first challenge involves the argument that his view does not give adequate expression to the importance of fundamental rights: In other words, the argument is that his proportionality account of constitutional rights renders them ‘too weak’. I shall seek to expand on Currie’s analysis here and largely agree with him that balancing does not necessarily weaken rights too significantly. However, I shall demonstrate that the objection does draw our attention to certain dangers attendant upon the balancing process when it is conducted by courts and highlights a central incompleteness in the balancing approach.

The second challenge I shall consider involves the argument that Alexy’s account of rights is ‘too strong’ and, if his approach is construed strictly, will render the limitation of rights very difficult to achieve. The discussion in this section will seek to consider the correctness or otherwise of Alexy’s characterisation of rights as ‘optimisation principles’. The discussion of both challenges highlights a number of important features that an adequate account of rights must take into account. In the last section of this paper, I attempt to use these insights to defend an alternative account of constitutional rights as being ‘weighted reasons protecting fundamental interests’. The last section will briefly seek to connect a structural account

4 Currie in Woolman & Bilchitz (n 3 above).
5 Woolman & Botha (n 1 above) 34-95 - 34-101.
6 In particular, Currie uses Alexy’s work on constitutional rights in R Alexy A theory of constitutional rights (trans Rivers) (2002).
of rights (such as that outlined by Alexy) with a more substantiative account of the nature of fundamental rights. In so doing, it will also seek to provide an understanding of at least one element of the substantive perspective from which a balancing enquiry should take place.

1 Does Alexy’s account of rights render them too weak?

1.1 Balancing and the weight of rights

Enshrining fundamental rights in a Constitution and in international treaties grants them a special status: It involves an assertion that such rights are higher-level norms with which other parts of the law and policy must conform.8 This in turn implies that such rights must have a particular importance for the individual or society such that they are enshrined above other norms. The idea must be that rights grant protections to individuals for interests that are of fundamental importance to them.9

This idea has been captured by a number of prominent legal philosophers in various ways. Dworkin, famously, has conceptualised rights as ‘trumps’ over utilitarian considerations.10 Though this is famously quoted as being Dworkin’s view, it is in some sense a misnomer as Dworkin does in fact recognise that collective interests may in certain circumstances prevail over rights where there are particularly strong reasons to do so.11 Importantly, though, for Dworkin is the fact that the very nature of rights means that we cannot override them simply on the basis of a routine calculation of costs and benefits.12 In a similar vein, Habermas conceives of rights as deontological ‘firewalls’, which insulate them against interference by ordinary reasons of policy or the general welfare alone.13

---

8 This occurs in what may be termed a dualist regime that ‘distinguishes … the higher law of the people from the ordinary law of legislative bodies’: J Rawls Political liberalism (1993) 233. See also F Michelman ‘Justice as fairness, legitimacy and the question of judicial review: A comment’ (2004) Fordham LRF 1412.
9 See, eg, D Bilchitz Poverty and fundamental rights (2007) 187. This is a substantive claim as distinct from the one made in analytical jurisprudence that the essential feature of all rights is that they protect interests: see MH Kramer et al A debate over rights (1998).
10 R Dworkin ‘Rights as trumps’ in J Waldron Theories of rights (1984) 153. See also R Dworkin Taking rights seriously (1977) 192 who states that ‘if a fundamental right exists, the government is not entitled to act ‘on no more than a judgment that its act is likely to produce, overall, a benefit to the community. That admission would make his claim of a right pointless, and would show him to be using some sense of “right” other than the strong sense necessary to give his claim the political importance it is normally taken to have.’
11 Dworkin (n 10 above) states, eg, at 191 that ‘[a]lthough citizens have a right to free speech, the government may override that right when necessary to protect the rights of others, to prevent a catastrophe, or even to obtain a clear and major public benefit’. D Meyerson ‘Why courts should not balance rights against the public interest’ (2007) Melbourne University LR 812 also makes this point.
12 Dworkin Taking rights seriously (n 10 above) 192.
Schauer similarly conceives of rights as ‘shields’. All these conceptions of rights seek to highlight the normative weight of fundamental rights in relation to other competing normative considerations and which explain their special place within the constitutional order.

Alexy puts forward a structural account of rights as ‘optimisation principles’. The problem posed by a number of theorists is whether his theory is capable of capturing the special normative weight of constitutional rights in relation to other competing considerations. It is necessary to outline briefly Alexy’s account of rights in order to engage with this criticism.

Rights, according to Alexy, are principles rather than rules. Rules are norms that are always either fulfilled or not; whereas principles are ‘norms which require that something be realised to the greatest extent possible given the legal and factual possibilities’. This characterisation of principles has implications for how to deal with conflicts between them: It means that where they conflict, one principle has to be weighed against the other. In this process, one of the principles may be outweighed by the other. Alexy contends that this process is governed by the principle of proportionality: Indeed, ‘the nature of principles implies the principle of proportionality’. As Currie points out, Alexy attempts to provide a rational justification for the way in which proportionality assessments are conducted in constitutional law as well as an analytical structure that renders the process less subjective.

Proportionality, Alexy claims, has a number of sub-principles: Two of these relate to more empirical matters. First, a particular measure that purports to restrict the realisation of a right must be closely connected to the realisation of another normative principle. This is the requirement of suitability: The South African Constitution incorporates this idea in section 36(1) by requiring that there be a close connection between the limitation of a right and the purpose it seeks to achieve. Secondly, flowing from the idea of rights as optimisation principles is also the requirement of necessity: This entails that if there are a number of measures that would equally give effect to a particular principle then the measure that would be least intrusive of competing normative principles must be chosen.

The other key element of proportionality in Alexy’s account relates to what he terms the realm of legal possibility: This involves the idea of

16 Currie has elaborated upon this in more depth in his paper and so I simply present a summary of the key points relevant to the argument here.
17 Alexy (n 6 above) 47.
18 Alexy (n 6 above) 50.
19 Alexy (n 6 above) 66.
balancing which is a process of assessing the ‘weight’ of the substantive considerations underlying each competing principle. As Alexy puts it, it is necessary to evaluate the ‘degree or intensity of non-satisfaction of, or detriment to, one principle versus the importance of satisfying the other.” 21 In seeking to achieve the optimisation of each principle and give each the normative weight they deserve, it is important that the following law of balancing be observed: ‘The greater the degree of non-satisfaction of, or detriment to, one principle, the greater must be the importance of satisfying the other.’ 22

It is important to recognise that balancing on this conception is not simply a mysterious process, subject to the play of intuition and lacking any rational basis; rather, it is indeed possible to judge objectively whether balancing has been performed correctly or not. As Kumm puts it, ‘[t]he metaphor of balancing should not obscure the fact that … the proportionality test will in many cases require the decision-maker to engage in theoretically informed practical reasoning, and not just in intuition-based classificatory labelling’. 23 I shall mention three possible ways in which to evaluate a court’s proportionality analysis in a particular case. First, it is important to consider the assignment of relative weights and intensities to various principles that the court is balancing: Has the court’s reasoning demonstrated that it assigns sufficient weight or intensity to the right and the competing considerations? The second area which admits of evaluation is the evidentiary basis the court uses for its evaluation and whether this has sufficient strength to support the assignment of particular weights or intensities to the various principles. The final evaluation one can make concerns a court’s comparative assessment of the relative principles and whether it has judged the matter correctly given the respective weighting of these principles. 24

1.2 Structure and substance: Do they complement one another in Alexy’s theory?

As Currie has outlined, the theory itself is impressive and is able to respond to a number of charges often made against limiting rights through a process of proportionality. But, the question that I shall seek to consider here is whether it adequately captures the nature of constitutional rights. Some critics claim that Alexy fails to capture one of the most essential features of such rights, namely, the fact that they have a special normative weight and

---

20 The South African Constitution also includes a provision that requires that ‘less restrictive means’ be chosen to achieve a purpose where this involves the limitation of rights. In the second part of this paper, I shall, however, seek to show that Alexy’s understanding of this requirement is too onerous and requires a revision in our understanding of rights as ‘optimisation principles’.

21 Alexy (n 6 above) 105.
22 Alexy (n 6 above) 102.
23 Kumm (n 15 above) 148.
24 I shall attempt to say a little more about some of the substantive considerations that should guide the court in the last section of this paper.
priority in reasoning related to constitutional law. This problem seems to flow from the fact that Alexy rejects the notion that the concept of a principle is to be tied closely to the notion of an individual right; rather, he seems to recognise that there are other types of normative considerations that also find expression in principles (or norms that can be optimised). No differentiation is made between principles that give expression to rights and those that give expression to other normative considerations. Where principles clash, ‘[c]ourts should assess the relative weight of these … and resolve the conflict in favour of the stronger reason’. The worry of these critics is that this process will devalue rights and lead simply to become one consideration to be weighed against many others; this process may also lead rights not to be accorded their proper weight in relation to other considerations. As Mattias Kumm points out, ‘[r]ights and policies compete on the same plane within the context of proportionality analysis’. This may lead, for instance, to situations where rights are overridden where they should not be as they are not accorded the weight they deserve in the proportionality analysis.

Meyerson, for instance, seeks to provide an alternative account of the way in which rights should be conceived in constitutional analysis. She distinguishes between making a decision on the balance of reasons – an approach she suggests is advocated by Alexy – and reasoning on a weighted balance of reasons.

When reasoning on the balance of reasons, decision makers assign all reasons the weight which they think they actually deserve. By contrast, when reasoning on a weighted balance of reasons, some … reasons are artificially assigned a different weight from the weight they would ordinarily receive … [S]uch an approach would acknowledge that bills of rights do not exclude consideration of the public interest, but would also build into rights adjudication a ‘systematic bias’ against permitting the infringement of rights.

Meyerson proceeds to offer arguments in favour of her approach which she claims is better able to capture the special priority that must be given to fundamental rights.

Meyerson’s charge against Alexy seems to be that the structural analysis of rights he proposes fails to give effect to the substantive normative weight that rights have. To evaluate whether Meyerson is correct or not, it is

25 Alexy (n 6 above) 66. This is contrary to the position of Ronald Dworkin who conceives of principles as norms ‘which can be offered as reasons for individual rights’ and contrasts them with norms which relate to collective interests (‘policies’): See Dworkin Taking rights seriously (n 10 above) 82 90.
26 Meyerson (n 11 above) 809.
27 Kumm (n 15 above) 142.
28 This is not merely a matter of theoretical concern and is a danger which seems to have occurred in several cases.
necessary to distinguish between structural and substantive accounts of rights. Alexy’s theory is essentially a structural one: It provides us with a formal analysis of the way in which rights reasoning takes place. His theory, however, whilst telling us that balancing is required when two or more principles are at issue, does not tell us anything about the actual substantive weight that is to be accorded to particular principles in the balancing process. The notion of balancing itself does not therefore determine the outcome in any particular case: for this, competing principles must be assigned weightings based upon their substantive importance.

Significantly, this means that the theory itself is fundamentally incomplete: It requires a substantive theory of rights and political morality in order to be operational. The very idea of ‘balancing’, and the ‘intensities’ and ‘weight’ of rights are mere metaphors which are dependent upon a substantive theory to give them content. As Alexy admits, ‘[p]roportionality analysis is … a formal structure that essentially depends on premises provided from outside’. It only makes sense to argue for the strong weight rights should have in our reasoning if we do recognise that they protect some of the most fundamental interests of individuals that deserve stringent protection. The formal analysis by Alexy is fundamentally dependent upon an understanding of the normative theories that provide a justification for fundamental rights. This point needs stressing, particularly in the context of the South African Constitutional Court, where there has often been an unwillingness to engage in substantive reasoning with regard to the normative content of rights and the countervailing considerations that are raised when seeking to limit rights. Formal reasoning is used to avoid and obscure the substantive evaluation that is required. The structure of balancing is no substitute

30 R Alexy ‘Thirteen responses’ in Pavlakos (n 15 above) 344. Some theorists also point out that there are certain reasons that should not be admitted into the balance at all: these are what Kumm refers to as 'excluded reasons'. Thus, arguing that same-sex marriage is unconstitutional because it does not accord with a particular person’s beliefs about what is Christian would not be a legitimate reason to enter into the balance. The justification for excluding such reasons must again be ones of substantive political morality though the point is important in recognising that balancing will only occur amongst ‘legitimate reasons’ and some reasons will have no force at all (we can capture this by saying that they may have zero weight). Others on occasion may come close to being conclusive without engaging in balancing (a prohibition on torture eg – though this is controversial in ticking bomb cases) and in such cases we may give them ‘infinite weight’ in the balancing process. These issues are complex and I cannot enter into a full discussion here: They raise questions concerning the nature of political liberalism as well as whether rights have a deontological element or not. See the discussion in Kumm (n 15 above) 131 - 166; see also Alexy’s reply in the same book at 340 - 344. D Meyerson Rights limited (1997) is partially concerned with outlining a theory of the kinds of reasons that may legitimately enter into balancing following a Rawlsian approach of ‘public reason’. The essence of Porat’s article (n 29 above) is to defend the notion that there are certain types of reasons that should not enter into balancing, a point similar to those made by the other authors above. A further discussion of this point lies beyond the scope of this piece.

for detailed engagement with the principles and values at stake in a particular instance. It can in fact only adequately be performed with an understanding of the content and justification of the normative considerations at stake.

When the very incompleteness of the balancing approach is recognised, it becomes possible to determine the contours that a defence of it can take. For, it clearly does not entail that every principle in the balancing process is to be accorded equal weight. Principles that give expression to rights can thus be weighted more heavily in the balancing process according to the importance that they hold as a matter of substantive political morality. As Kumm puts it, ‘[t]he fact that the proportionality analysis does not prioritise individual rights over collective goods on the structural level, then, does not mean that such a priority cannot be given adequate expression within that structure’.

Alexy’s theory thus can allow for giving special weight to rights in constitutional law given that they are principles that are of particular importance in a political community. If the correct weight is assigned to rights, however, there would seem to be no important difference between reasoning on a balance of reasons and reasoning on a weighted balance of reasons. The problem for theorists such as Meyerson is that the moment they admit that rights do not have absolute weight in respect of all competing principles (a claim they do not wish to make), they admit the possibility of limiting rights under certain circumstances. In determining these circumstances, there will have to be an assessment of countervailing principles and, consequently, the need to engage in an assessment of the relative normative weight of these principles: ie balancing. As long as rights are given the correct weighting, it does not seem that the wrong results will be achieved.

---

33 Kumm (n 15 above) 149.
34 It also allows for assigning different weights to different rights depending on their relative importance in a particular case.
35 Meyerson attempts to set up a second-order level of reasons (following a suggestion of Raz) that generally trump the first-order level of reasons. Rights would essentially be second-order reasons that generally take precedence over other considerations at the first level. Yet, the problem lies in the word ‘generally’. If rights always trump first-order reasons, then they are absolute with respect to such reasons and we will need a good case to show why this is so. If, however, the notion of absolute trumps is not convincing (as Dworkin, eg, admits and most theorists agree), then we will always have to consider the relationship between the one set of reasons and the other. The attempt to insulate one set of reasons from being balanced in relation to another set is not convincing unless one has already engaged in substantive reasoning that establishes the relationship between the two sets of reasons. Rights will have the weight they are to be attributed as a matter of substantive morality in relation to other considerations. If no absolute priority is to be established, then there will have to be a form of balancing involved in setting up this relationship between different types of reasons in the first place. I considered at n 31 above the challenge that certain sets of considerations should not be placed in the balance at all.
1.3 Should we over-protect rights?

If the balancing process can give adequate expression to the weight of rights, then it seems it is able to capture the normative strength of rights. However, in response, theorists such as Meyerson could claim that their problem with Alexy’s theory is its failure to include a presumption in constitutional law in favour of rights that can only be dislodged by strong countervailing reasons. Such a presumption can be justified by an argument concerning the institutional difficulties of accurate balancing: The fact that the assessment of the relative weight of principles is an ‘uncertain, complex and speculative matter as attested by the frequent disagreements among judges on these questions’. In certain circumstances, judges might be ‘subjectively predisposed to underestimate the strength of rights when they come into conflict with government goals, especially in areas like preventing crime and protecting national security’. Given the likelihood of forms of institutional bias in certain cases, it may be justifiable to ‘over-protect rights’. This could be achieved, for instance, by requiring that those seeking to limit rights need to defeat a legal presumption in favour of those rights. Alternatively, it might be that the justification for limiting rights must be shown to meet a higher constitutional standard such as realising a ‘compelling interest’.

The example given by Kumm illustrates this point well: A case came before the European Court of Human Rights as to whether it is discriminatory to exclude lesbian and gay people from the military. Kumm recognises that naked homophobic reasons for excluding lesbian and gay people from the military should not be considered (or given zero weight in the balancing process). However, he worries that arguments concerning the impact of their inclusion upon the fighting power and operational effectiveness of the military may weigh heavily with courts in any balancing procedure. Courts may land up sanctioning homophobic attitudes through an institutional bias in favour of granting the military a discretion over matters of defence and national security. To counteract such an institutional bias, such theorists claim it is necessary to ‘over-protect rights’ and thus create a strong presumption in their favour which it is difficult to overturn.

It should be recognised here that this objection does not attack the appropriateness of balancing in determining the outcome that should be achieved in a particular case. Even with the procedural correctives suggested, some form of balancing will still have to occur though the dice

---

36 Indeed, Schauer (n 14 above) 429 - 430 claims just that: Decision makers should see rights as shields that presumptively but not conclusively take precedence over other governmental interests.
37 Meyerson (n 11 above) 817.
38 As above.
39 This idea is suggested by Kumm (n 15 above) 152.
40 Lustig-Prean and Beckett v United Kingdom [1999] ECHR 71 (31417/96; 32377/96, 27 September 1999). The case is discussed in Kumm (n 15 above) 152.
will be loaded in favour of rights. What these theorists highlight is rather a concern about the likelihood that courts (and possibly other institutions) will not conduct the process of balancing properly and so fail to give rights the weight they deserve in such a process.

The critique of Alexy thus highlights the importance of ensuring that the structure of rights analysis is congruent with its purpose. I would suggest that congruency is not sufficient: The form should itself give expression to the purpose of rights protection. The worries of Alexy’s critics in relation to the form of his theory should not be easily dismissed. The evaluation of rights against other considerations on the same ‘level’ may tend to lead courts to fail to give sufficient strength to rights and thus lead to their under-protection. It may thus be necessary to develop procedural safeguards that ensure rights are given the strength they deserve in any constitutional analysis. The two-stage procedure in South African constitutional law assists in ensuring that at least courts begin by thinking through the nature of the violation that occurs in relation to a particular constitutional right. It is thus critical that courts do not omit the first stage of the analysis and seek to capture the interests and values at stake in a particular violation. The second stage of the inquiry will involve a proportionality analysis which must take account of the ‘nature of the right’: This factor will require understanding the strength and intensity to be given to a particular right. 42

Yet, it might be thought that these procedural elements remain insufficient to deal with the likelihood of error on the part of courts in giving rights the strength they deserve. Various possibilities suggest themselves: One has already been canvassed in the form of a presumption in favour of rights. Alexy suggests another possibility for dealing with the problems identified by Kumm that would not systematically over-protect rights but, he claims, deal with the likelihood of error in such cases.

He indicates that these problems arise particularly when empirical premises lie at the foundation of a claim to limit rights: the proposition, for instance, that the inclusion of lesbian and gay people in the military impacts upon its fighting power and effectiveness. Alexy consequently proposes a second law of balancing to be included within the proportionality analysis: ‘The more heavily an interference in a constitutional right weighs, the

---

41 I thus support the arguments of Woolman & Botha (n 1 above) 34-18 - 34-29 where they criticise certain *dicta* by Justice Sachs that suggest a possible conflation between the two stages of the inquiry. They also correctly criticised decisions such as *Christian Education* where the first stage is virtually entirely omitted. For a recent example of the majority of the Court failing adequately to conduct the first stage of the analysis, see *Road Accident Fund v Mdeyide* 2011 2 SA 26 (CC) and D Bilchitz ’How should rights be limited? *Road Accident Fund v Mdeyide* (2011) Journal of South African Law 568.

42 In the last section, I will suggest why, when we understand the nature of the substantive enquiry we are to conduct, there would also be a systematic bias in favour of rights protection.
greater must be the certainty of its underlying premises.\textsuperscript{43} This principle would require courts only to accept relatively conclusive evidence on the basis of which to limit a fundamental right: speculation and weak causal or evidentiary claims would not be accepted.\textsuperscript{44}

Alexy’s point here seems to be well-taken and is important for courts to bear in mind when conducting a proportionality analysis. It is questionable, however, whether it provides a fully adequate response to the concerns of theorists such as Kumm and Meyerson: The problems they fear seem capable of arising in the assessment of the relative weight of principles in the balancing process even where empirical questions are not at issue. Importantly, I have sought to argue in this section that their critique of Alexy’s theory does not fundamentally impugn his conceptualisation of rights or of proportionality; what it does do is place the focus on certain features of his theory that are both incomplete and pose dangers for the protection of rights through courts. The logical conclusion would be to adopt certain procedural safeguards in legal systems using proportionality analysis that could correct for the under-protection of rights towards which it may lead. There is also a need to adopt a substantive conception of rights that links to the structural one. I now turn to consider a critique of Alexy’s conception of rights that draws out some of the implications of his view to suggest that far from offering ‘too weak’ a conception of rights, his understanding of them may be ‘too strong’.

2 Does Alexy’s account of rights render them too strong?

This objection is rooted in Alexy’s conception of rights which, as has been seen already, he characterises as principles rather than rules. Dworkin gives a relatively modest account of principles as stating ‘a reason that argues in one direction, but does not necessitate a particular decision’.\textsuperscript{45} He also contends that an integral part of the concept of a principle is that it has the dimension of ‘weight’, which gives expression to its importance in any particular setting.\textsuperscript{46} Alexy seeks to build on Dworkin’s account of principles (and the distinction between them and rules) but goes one step further in characterising principles as optimisation requirements, norms that must be fulfilled to the ‘greatest extent possible’.\textsuperscript{47}

\textsuperscript{43} Alexy (n 30 above) 342.
\textsuperscript{44} Justice Ngcobo, writing for the minority, made a remark that indicates his acceptance of this point in Prince v President, Cape Law Society 2002 2 SA 794 (CC) para 74 where he stated that ‘[i]n my view, a constitutional right cannot be denied on the basis of mere speculation unsupported by conclusive and convincing evidence’.
\textsuperscript{45} Dworkin (n 10 above) 26.
\textsuperscript{46} Dworkin (n 10 above) 27.
\textsuperscript{47} Alexy (n 6 above) 47. Alexy (in n 27 of ch 3 at 48) notes that he differs from Dworkin on the characterisation of principles as optimisation requirements.
The latter claim could be broken down into the following elements: First, in the absence of a competing principle, the principle requires complete realisation. If we have a right to privacy implicated in a particular case and no competing principle, then that right must be realised in those particular circumstances. Secondly, where there are competing principles, the principle of proportionality will determine the outcome (with its sub-principles of necessity, suitability and balancing): Ultimately, though, the notion of principles as optimisation requirements means that the outcome must be adopted that allows for the greatest possible realisation of each principle so far as this is compatible (both factually and legally) with the greatest realisation of the others.48

It is important to recognise that Alexy’s language is metaphorical. The notion of ‘optimisation’ in a field such as economics, where clear quantification is possible, has a clear meaning: We clearly understand what it means, for instance, to optimise profit. However, the exact meaning of this term in the legal or moral context is less clear as the principles and values to be evaluated and balanced are not clearly quantifiable in the same manner.49

A crucial notion in understanding what optimisation means in the legal context is the idea of ‘possibility’.50 When referring to what is legally possible, Alexy seems to be employing the idea of logical possibility.51 However, it remains unclear exactly what he means by the idea of factual possibility: Does this refer only to the outer limits of physical possibility (a measure cannot involve requiring humans to fly as they are not able to do so without being aided by an aircraft) or does this notion include constraints such as economic scarcity (a measure is unaffordable for a country given current lending policies, although it could be affordable if these policies changed) or political sensitivity (a measure cannot be passed as it will cause massive protests and civil disobedience if it is)? Determining what is included within the realm of factual possibility will be critical for determining the extent to which a principle can be realised or not and thus be a crucial factor in the application of the principle of proportionality. Without clarifying what is meant by possibility, the balancing process may not lead to determinate outcomes as it will not be clear what is meant by realising a principle to the ‘greatest extent possible’. Moreover, if the account of possibility that Alexy gives is too strong (such as conceiving factual possibility as the outer limits of physical possibility), then it would render it extremely difficult to justify the limitation of fundamental rights even in cases where this seems intuitively to be reasonable.

49 Moller (n 48 above) 462.
50 Moller (n 48 above) 459.
51 Alexy (n 6 above) 51 - 53.
The lack of clarity on the nature of the notion of possibility employed in Alexy’s theory is linked to a more general problem involved in conceiving of rights as requirements that must be realised to the greatest extent possible. I shall now proceed to consider a key implication of this claim: the requirement that the means adopted to limit a right must be the least restrictive of the right in comparison to other possible alternatives (the so-called ‘necessity’ component of proportionality).

2.1 The ‘necessity’ problem

The principle of necessity requires ‘that of two broadly equally suitable means, the one which interferes less intensively [with a right] should be chosen’.52 If rights are norms that must be optimised to the greatest extent possible, then when engaged in proportionality analysis, courts should adopt a stringent approach ensuring that any limitation of a right that is being proposed should strictly speaking be necessary to realise the purpose sought to be achieved. This means that no other alternative must be available that can equally realise the purpose and be less invasive of the right in question. Let us refer to this as the strong interpretation of necessity. According to the logical derivation of the principle of necessity given by Alexy, the strong interpretation is entailed by an understanding of rights as optimisation requirements.53

The question that arises is whether such a strong interpretation of necessity is desirable when conducting a proportionality enquiry. An example given by Alexy can help to illuminate this point. The Federal Constitutional Court in Germany was presented with a case in which they had to decide whether or not a government-imposed requirement that tobacco manufacturers had to place health warnings on cigarette packets was unconstitutional or not.54 The tobacco manufacturers claimed that their right to trade freely had been unjustifiably infringed; the government argued that these health warnings were of importance in helping to deter smoking and to protect individuals against the severe health risks associated with it. Alexy analyses the case as follows: He contends that the government measure involves a relatively minor interference with the freedom of trade of tobacco manufacturers. On the other hand, the impact of smoking on human health is severe, leading to cancer and cardiovascular disease. Consequently, in the proportionality analysis, we have to compare a relatively minor interference with a right against a very strong case in favour of the limitation in question. Consequently, he concludes that the limitation on the rights of tobacco manufacturers, in such an instance, would be justified and the law in question would pass constitutional muster.

---

52 Alexy (n 6 above) 398.
53 Alexy (n 6 above) 67 - 68, 399.
54 BVerfGE 95, 173. Alexy (n 6 above) discusses the case at 402.
Whilst Alexy is persuasive in his analysis of the balancing process at issue (the question of legal possibility), he does not deal with whether the government measure in question was in fact necessary for achieving the purpose in question. If we take the strong view of necessity, then a measure will only be justifiable if there are no less intrusive means available to realise the goal in question. One of the really difficult problems flowing from Alexy’s theory lies in evaluating the claim made by the legislature in this instance that there are no alternative means available other than the placement of health warnings on cigarette packets that are less restrictive of the rights of the manufacturers and would equally serve to achieve the purpose in question.

Indeed, when we begin evaluating this case, the uncertainty of any necessity claim (and in fact the suitability claim as well) made by the legislature becomes more evident. First, establishing an empirical grounding for its claim would require showing that placing health warnings upon cigarette packets are likely to result in a decrease in smoking: The difficulties of producing such empirical evidence should be evident (this would thus challenge the empirical grounding for the ‘suitability requirement’). Secondly, alternative measures could easily be developed which do not require tobacco producers to place health warnings on their packets (and thus be less restrictive of the right in question). Indeed, one may question the very effectiveness of such a government measure given that the people buying the cigarettes are likely to be the ones least interested in these health warnings. Thirdly, broader public educational measures, or a major advertising campaign, would perhaps be more effective means to achieve the public health objectives set by the government and be less restrictive of the rights of manufacturers.

Even if a court were to disregard these worries, and find that placing warnings on cigarette packets was a necessary measure to achieve the government’s legitimate aims, it is also important to recognise that the manner in which such a measure is implemented will clearly be relevant to a determination of its proportionality and whether it is least restrictive of the manufacturers’ rights. It would be possible, for instance, for the government to subsidise any costs incurred in the process of placing these health warnings which would seem to be less restrictive of the commercial interests of the tobacco manufacturers. Should courts therefore require governments to pay for such measures where rights are implicated? Moreover, it seems that there would always be a very difficult determination to be made by courts as to the acceptable size of any such health warning: was the size of the warning required by the legislature the least intrusive possible of the manufacturers’ right to place commercial information on the cigarette packet?

The questions I have raised are not designed to demonstrate that the case in question was wrongly decided in favour of the government. Indeed, it seems intuitively clear to me that there is a reasonable case to be made that health warnings on cigarette packets represent a justifiable interference with the rights of commercial manufacturers in the interests of public
health. The problem, however, lies in Alexy's conception of rights as optimisation requirements: for it entails that any measure that limits a right must strictly speaking be such that there are no alternative measures that could be capable of realising the purpose of the limitation and be less intrusive of the right. A sensible measure such as the one the legislature enacted in this case could easily fail to pass the strict test required by this element of Alexy's proportionality analysis. Consequently, it would seem that his theory can place unreasonable demands on those seeking to justify the limitation of rights.

Indeed, the problem seems to be capable of being generalised: For, it appears likely that, in many instances, there will be a range of policy alternatives that may arguably intrude to a lesser degree upon a right than the measure adopted by the legislature in any particular case. This will be so particularly if the notion of possibility in Alexy's theory is understood to include all measures that are within the realm of physical possibility. Several issues arise in this regard. The first difficulty lies in making judgments concerning the extent of interference of different measures. Even if we accept Alexy's point that we can roughly make such judgments, the second problem lies in the difficulty (if not impossibility in many instances) of proving that a particular measure is clearly the least intrusive of the right in question given the range of measures that could be adopted. It needs to be stressed that the latter requirement is not simply a claim that the measure in question is one of a range of measures that does not too severely infringe upon a particular right: it is rather the claim that this measure is the only one available of a range of alternatives that limits the infringement of the right to the greatest extent possible. A third problem flows from the others: In a constitutional review system, it will require courts to engage in an evaluation of competing alternative policy measures which could strongly intrude into the domain of the legislature. The very stringency of the enquiry entailed by Alexy's theory could thus systematically breach the respective roles of courts and other branches of government. This has clearly been a worry evident in the reasoning of the South African Constitutional Court when engaging with the 'less restrictive means' element of the limitations enquiry. I now turn to discuss the jurisprudence of the Court in this regard and shall argue that it has displayed an acute awareness of the problems that could be caused by a stringent interpretation of the necessity component in the proportionality enquiry.

2.2 SA Constitutional Court’s approach to ‘necessity’

As has been mentioned, the South African Constitution contains a general limitation clause which outlines the circumstances under which rights may
be limited. The clause requires a broad assessment to be made concerning whether a limitation is reasonable and justifiable within an open and democratic society based on the values of human dignity, equality and freedom. It also outlines a range of specific factors the court must engage when determining the reasonableness of a limitation. These have generally been regarded as involving the broad components of a proportionality enquiry. These factors include the requirement that the limitation must be related to the purpose thereof (Alexy’s suitability requirement); and also a consideration by courts as to whether there are ‘less restrictive means to achieve the purpose’ in question. The question arises whether the latter factor involves Alexy’s ‘necessity principle’. Importantly, the wording of the Constitution refers to ‘less’ restrictive means rather than the ‘least’ restrictive means: the latter wording would be more consistent with Alexy’s necessity principle. Importantly, the way in which the Constitutional Court has applied this factor also suggests its interpretation thereof is less onerous than that proposed by Alexy.

A particularly important case in this regard is *S v Manamela*. The case concerned the constitutionality of a reverse onus provision that required an accused person who was in possession of stolen goods to prove that he or she had reasonable cause at the time of acquiring the goods to believe that they were not stolen. If the accused failed to discharge the onus, he or she would be guilty of the offence of receiving stolen property knowing it to be stolen. The majority found that the provision was in fact unconstitutional: Part of its reasoning in reaching this conclusion was that the statute could have placed an evidentiary burden on the accused rather than a full onus and that this would have been a less restrictive infringement of the rights of the accused. The minority (O’Regan J and Cameron J) disagreed with this position, finding that the imposition simply of an evidentiary burden would ‘diminish the obligation upon members of the public to act vigilantly’.

Despite the differing conclusions reached, both the majority and minority agreed on the approach to be adopted towards the less restrictive means requirement (the disagreement lay in the application of this approach to the particular case at hand). The minority judgment contains the main discussion in this regard which commences with the recognition that the ‘less restrictive

55 Sec 36(1) reads as follows: ‘The rights in the Bill of Rights may be limited only in terms of a law of general application to the extent that the limitation is reasonable and justifiable in an open and democratic society based on human dignity, equality and freedom, taking into account all relevant factors including (a) the nature of the right; (b) the importance of the purpose of the limitation; (c) the nature and extent of the limitation; (d) the relation between the limitation and its purpose; and (e) less restrictive means to achieve the purpose.’

56 See *S v Makwanyane* (n 1 above) para 104.

57 See *S v Makwanyane* (n 1 above) para 104.

58 *S v Makwanyane* (n 1 above) para 104.

59 *Manamela* (n 57 above) para 97.
Does balancing adequately capture the nature of rights?

Does balancing adequately capture the nature of rights? 283

means' component is an 'important part of the limitation analysis' but that it is 'only one consideration relevant to that analysis'. This factor requires 'a careful analysis of the purpose of the provision' that is being impugned. The judgment then goes on to quote Blackmun J in the US who famously wrote that '[a] judge would be unimaginative indeed if he could not come up with something a little less “drastic” or a little less “restrictive” in almost any situation, and thereby enable himself to vote to strike the legislation down. This argument leads the judgment to recognise that the legislature adopts the particular measures it does in light of 'concerns relating to cost, practical implementation, the prioritisation of certain social demands and needs and the need to reconcile conflicting interests'. The legislature, the judges hold, is the appropriate institution to make difficult policy choices and the court should 'take care to avoid a result that annihilates the range of choice available to the legislature. In particular, it should take care not to dictate to the legislature unless it is satisfied that the mechanism chosen by the legislature is incompatible with the Constitution'.

The latter claim of course does not help solve the problem as the question is when a particular measure limiting a right is in fact incompatible with the Constitution, and the less restrictive means requirement is meant to assist in determining this issue. Woolman and Botha, however, capture the difficulty being gestured towards by the judges in this passage when they understand its concern to mean that 'less restrictive means can often be envisaged, but such means may impose significant administrative burdens on the state, have other substantial cost implications, undermine the state's symbolic objectives or reverse a hierarchy of needs worked out through the political process'. The difficulty lies in determining when a court should strike down a measure where less restrictive means exist despite the fact that there would be costs to the political community in doing so. The way in which the Constitutional Court has sought to resolve this problem (although not in an entirely clear and satisfactory manner) can be gleaned through the following important dictum of Kriegler J:

Where section 36(1)(e) speaks of less restrictive means it does not postulate an unattainable norm of perfection. The standard is reasonableness. And in any event, in theory less restrictive means can almost invariably be imagined without necessarily precluding a finding of justification under the section. It is but one of the enumerated considerations which have to be weighed in conjunction with one another, and with any others that may be relevant.

60 Manamela (n 57 above) para 94.
61 Manamela (n 57 above) para 96.
62 Manamela (n 57 above); Illinois State Board of Elections v Socialist Workers Party et al 440 US 173 (1979) 188 - 189
63 Manamela (n 57 above) para 95.
64 As above.
65 Woolman & Botha (n 1 above) 34 - 89.
66 S v Mamabolo 2001 3 SA 409 (CC) para 49.
Chapter 12

This brief analysis of certain decisions of the Constitutional Court already enables us to conclude that its approach to the less restrictive means requirement does not accord with the strict ‘necessity’ approach that is entailed by Alexy’s conceptualisation of rights as optimisation requirements. A particular provision which limits a right will be able to pass constitutional scrutiny even if it is not, strictly speaking, the least intrusive measure that can be conceived or that lies within the range of physical possibility. The court rather can be understood to adopt a less rigid interpretation of the ‘less restrictive means’ requirement: Its interpretation rather requires that the means adopted must not too severely intrude upon a right in relation to other readily available alternative means that would equally realise the objectives sought to be achieved.

In other words, the less restrictive means requirement can be understood to involve several assessments: First, a court must consider the degree of interference that the means adopted by a legislative scheme causes to the right in question. Secondly, there needs to be consideration given to alternative means that could be adopted to realise the objectives of the limiting provision, though such an assessment need not exhaust the field but only engage with those that are clearly relevant and can be practically realised. Finally, there needs to be an overall assessment of the extent to which each of the measures considered interferes with the right in question and whether the one adopted by the legislature too severely intrudes upon the right in relation to the readily available alternatives. Critically important here is to recognise that the limitations clause does require a comparative analysis of certain ‘leading alternatives’ to the impugned measure. The question that must be determined, however, is not whether the ‘best’ measure has been adopted that is least restrictive of the right; rather, it is whether the one that has been adopted meets a threshold qualitative criterion of not being ‘too severely restrictive of a right’ in relation to the leading alternatives. It is possible on this interpretation of the requirement for there to be a measure that is less restrictive of a right than the one that is impugned (and passed by the legislature) but for a court nevertheless to find that the legislative measure is not ‘too severely restrictive’ of the right such that it still passes constitutional muster. Thus, at the heart of this interpretation of the Constitutional Court’s understanding of this component, lies a qualitative judgment concerning the degree of interference with a right by a particular measure in relation to readily available alternatives.

I have thus sought to show that Alexy’s conception of rights as optimisation requirements entails an overly onerous ‘necessity principle’ that would only allow for the limitation of a right to be justifiable in circumstances where such a limitation is the least restrictive measure that could be adopted. I have sought to show the major difficulties this would raise for a party seeking to justify the limitation of a right, thus leading to very few circumstances in which limitations would pass constitutional muster. I have also sought to demonstrate the fact that the South African Constitutional Court has not adopted such a stringent approach and is mindful of a number
of the difficulties faced by providing such a stringent interpretation of the circumstances in which rights may be limited. If, however, the reasoning thus far is correct, and the more moderate approach concerning 'less restrictive means' is preferable, then this would seem to pose a challenge to conceptualising rights in the way Alexy does – namely, as optimisation requirements. For, the principle of 'necessity' appears to flow logically from this conception. If Alexy's understanding of rights is flawed and yet the idea of proportionality and balancing remains important to the limitation of rights, what conception of rights should be put in its place? The concluding section of this piece seeks to provide some understanding of the contours of a more adequate conception of rights.

3 Conclusion: Towards an alternative view of rights and limitations

The first part of this article questioned whether Alexy's theory gave adequate expression to the special normative role of rights in constitutional law reasoning. I have sought to recognise the importance of balancing and proportionality: a requirement that flows from the fact that there are competing normative considerations in our social order. This discussion drew attention to the fact that the proportionality enquiry advocated by Alexy could attach special weight to rights considerations. Nevertheless, it highlighted the fact that a balancing enquiry may lead to the undervaluation of rights that might result from structurally placing rights on the same level as other types of normative considerations. Both procedural and substantive measures are necessary to counteract this tendency. In particular, it was stressed that it is important to recognise the relationship between the structural aspect of rights and their substantive justification and content.67

The second part of this article, however, questioned whether a conceptualisation of rights as optimisation requirements does not imply too strong a principle of necessity that places overly stringent constraints on those seeking to limit rights. For this would seem to imply that individuals would have to show that no other measure than the one impugned could achieve the desired objective and intrude on rights to a lesser degree.

Understanding the significant issues underlying these two criticisms allows us to begin to construct an alternative account of how we should understand rights in our constitutional scheme.68 In this process, it is important to return to Dworkin's conception of a principle as stating ‘a

67 Moller (n 48 above) 468 also recognises the insufficiency of a purely structural account of rights and the need to include a substantive dimension.
68 This section will seek to draw together some of the insights gained through the discussion of Alexy's theory. The constraints of the article render it possible only to begin on the path of constructing an alternative account of the nature of rights.
reason that argues in one direction, but does not necessitate a particular decision’.69 This conception of a principle is similar to what Jonathan Dancy has recently called a contributory reason: ‘This is a feature whose presence makes something of a case for acting, but in such a way that the overall case for doing that action can be improved or strengthened by the addition of a second feature playing a similar role’.70 Such a reason can also be defeated by another contributory reason that makes a case for a different course of action.

The pure notion of a principle or reason discussed by these authors is not sufficient to provide us with an adequate account of rights.71 Two crucial elements have been left out, the one formal and the other substantive. The first element relates to the fact that we started with, namely, that rights are not just ordinary principles or reasons: They have a special character in our constitutional scheme and are deserving of special protection. Thus, we cannot simply regard them as principles: they are a special type of principle that possesses weightiness in our moral and political reasoning. I shall thus seek to indicate this element of rights by referring to them as ‘weighty principles’. Importantly, this conception of weighty principles does not require that a principle be understood to be an optimisation requirement: Rather, what is critical is that in our reasoning it be given the appropriate weight that is to be attached to it. This formal conception of rights helps us to make sense of the moderate interpretation of the Constitutional Court’s approach to the less restrictive means requirement in the limitation clause. Instead of requiring that the impugned measure be the least restrictive means in relation to competing measures, the Court simply requires that appropriate weight be given to the right when differing and other possible measures are under consideration.

The second element seeks to capture the underlying reason for rights having this special character of being ‘weighty principles’: They have a particular substantive content and function in our political system. One of the problems with Alexy’s theory is that his account of rights could apply to any normative consideration that requires optimisation to the greatest possible extent. The theory seeks to detach a formal essence of rights from their substantive content. It is important, in a revised conception of rights, that we fuse the two elements closer together to ensure that form does not become an end in itself and that it rather follows its function. To develop an adequate conception of fundamental rights, it is thus necessary to have an understanding of their substantive justification, role and purpose within

---

69 Dworkin (n 10 above) 26.
71 I have some anxiety even about referring to rights as principles or reasons which seems to denude them of their very essence. They appear rather to be a type of claim that a being has by virtue of certain properties of its nature to be accorded certain types of treatment. It is not clear that the notion of a claim here is identical to the notion of a principle as described by Dworkin, Dancy or Alexy. A further discussion of this issue, however, lies beyond the scope of this paper.
a constitutional scheme. Whilst I cannot seek to accomplish this in an exhaustive manner here, I would suggest that as a minimal starting point, it is of importance to recognise that constitutional rights seek to protect some of the most fundamental interests of individuals. This feature of rights helps to explain their weight. They relate to the very most basic elements of our lives – the necessary conditions of our freedom, the resources we need to live lives of value, and to function adequately. Rights have a form of urgency that flows from the impact that they have on our ability as individuals to lead lives of value to us.

Rights are thus best understood as ‘weighty principles that protect the most fundamental interests of individuals’. At their foundation is a concern for the value and quality of individual lives. In a just political community, each individual life will be understood as having equal value. Thus, rights can be understood to flow from what may be termed the ‘equal importance of individual lives’. This understanding of rights is fundamentally congruent with the text of the South African Constitution. The fundamental values underlying the Constitution are human dignity, equality and freedom. Human dignity involves a fundamental assumption of the worth of each individual life and thus captures the notion of the importance of individual lives. Equality represents the notion that each life is to be treated as equally valuable. Freedom is part of what constitutes the value in individual lives: to be able to flourish and live according to one’s own conception of the good.

Importantly, these three foundational values are also the ones that play a part in determining whether the limitation of a right is justifiable or not. This is a crucial point as it places at the heart of the enquiry to be conducted under the limitations clause the very values that underlie fundamental rights. This means that one of the key normative elements when deciding whether a

---

72 Dworkin (n 10 above) attempts to capture the importance of the interests underlying constitutional rights in his discussion at 191 - 192; M Nussbaum Women and human development (2000) 96 - 101 also provides an account of rights as essentially protecting important interests (or in her terms capabilities) of individuals.

73 See Bilchitz (n 9 above) 187: This fact may mean that there may be different elements within a particular right that have greater priority and urgency than others. This is of particular importance in the context of socio-economic rights: See Bilchitz (n 9 above) 185 - 194. This would help explain why at times certain interferences with rights may be more objectionable than other interferences.

74 Dworkin regards the notion of equal concern and respect to be accorded to individual lives as the sovereign virtue of a political community: See R Dworkin Sovereign virtue (2000) 1. I have sought to work out this idea in relation to fundamental rights in Bilchitz (n 9 above) 57 - 65.

75 See Bilchitz (n 9 above) 57 - 65.

76 Sec 1 of the Constitution.

77 See Dawood v Minister of Home Affairs; Shalabi v Minister of Home Affairs; Thomas v Minister of Home Affairs 2000 3 SA 936 (CC) para 35. In D Bilchitz 'Does transformative constitutionalism require the recognition of animal rights’ in S Woolman & D Bilchitz (eds) Is this seat taken? Conversations at the Bar, the bench and the academy about the South African Constitution (2012) 173, I argue that dignity should not be confined to human beings alone.
limitation is justifiable or not is whether it adequately takes into account the ‘equal importance of individuals’. This means that, already integrated into the limitations enquiry itself is a systematic bias in favour of rights: For the very considerations underlying the recognition of rights in the first place are a crucial part of determining whether a limitation will itself be justifiable or not. In a sense, this helps meet the concern raised earlier in the discussion of Meyerson’s objection: The very perspective from which the proportionality enquiry is to be conducted is friendly towards rights and thus is likely to give substantive weight to them.

This understanding of the limitations enquiry also offers a response to the critique of Botha and Woolman to the effect that balancing requires the evaluation of incommensurable principles. Currie counters this by arguing that in fact, despite the difficulty, constitutional principles can often be compared and balanced: they are not in fact incommensurable. He claims that they are to be evaluated against the ‘perspective of the Constitution’.79 I agree with Currie that claims made by those in favour of incommensurability are problematic (and possibly incoherent) and that it is necessary to have a perspective from which competing principles will be evaluated.80 It is important, however, to say a little more about what the perspective of the Constitution entails if this idea is not to be wholly indeterminate. The analysis provided here neatly complements Currie’s response in providing greater content to what the perspective of the Constitution is: It specifies that the relevant perspective can be drawn from the three values contained in the limitations clause – dignity, equality and freedom. In other words, when evaluating competing principles, we consider them, at least partially, in relation to the extent to which they are capable of realising the ‘equal importance of individuals’. When faced, for instance, with a conflict between privacy and security, it is important to recognise that they both have an impact on individual lives. Though they may seem difficult to evaluate against one another, where they conflict, the necessary perspective that must be adopted in evaluating any particular limitation concerns the respective impact of particular measures upon the equal importance of individuals.

I do not seek to suggest that ‘equal importance’ is the only factor that courts should consider in the limitations enquiry but it is one of central significance. It allows us to understand the substantive focus upon individuals and the quality of their lives that is at the heart of the Constitution. It also assists us in judging competing principles and places the values underlying rights at the very heart of the judgment that has to be made when deciding whether a particular limitation is justifiable or not.

79 Currie in Woolman & Bilchitz (n 3 above).
80 See VA da Silva ‘Comparing the incommensurable: Constitutional principles, balancing, and the rational decision’ (2011) Oxford Journal of Legal Studies (forthcoming) for the argument that whilst constitutional principles may be incommensurable they are comparable.
I have thus sought in this article to show that understanding the way in which rights can be limited is not just a technical matter of constitutional law doctrine. It goes to the heart of what it is to have a fundamental right. It is imperative that the courts in their reasoning about rights understand what they are doing and so ensure that rights are not treated as simply one consideration to be balanced against many others, but are given the normative weight they deserve. Such weight should not, though, preclude the limitation of rights where this is, in fact, warranted. Balancing indeed has an important part to play in constitutional law: it must, however, be supplemented with a proper structural and substantive understanding of rights.
1 Introduction

This article endeavours to construct a framework for understanding the concept of constitutional deference. Much has been written about deference in the past two decades, and this article attempts to piece together the skeleton for a better understanding and the beginning of a dialogue over the kinds of values which ought to inform the South African courts’ approach to constitutional deference. What this article is not is a normative argument about the courts’ approach to executive and legislative decision making. Such an argument needs to be premised upon a particular conception of democracy and the state – an approach which is not uncontested in the text of the Constitution\(^1\) – and which is beyond the scope of what is attempted here.

Instead, what this article does seek to do is to provide the tools for a dialogue for what is termed ‘constitutional deference’, that is, an approach to deference rooted in our constitutional democracy and our courts’ approach to adjudication. The article thus seeks to create a framework for an understanding of constitutional deference which will illuminate the courts’ approach to judicial decision making. While the framework is intended to be applicable to the adjudication of all disputes, this article selects the adjudication of socio-economic rights by the Constitutional Court as an example of how this framework could usefully illuminate the approach of the courts to executive and legislative decision-making.

\(^*\) I would like to thank Cora Hoexter, Redson Kapindu and the anonymous referees for comments on earlier drafts of this article. This article is derived in large part from the book K McLean *Constitutional deference, courts and socio-economic rights in South Africa* (2009).

\(^1\) Constitution of the Republic of South Africa, 1996.
The article is structured in two main parts. The first part provides a framework for the understanding of constitutional deference as consisting of three, sometimes competing, considerations which underpin a particular court’s approach. In the second part, this framework is applied to an examination of the adjudication of socio-economic rights by the Constitutional Court.

2 Framework for understanding constitutional deference

There are a number of reasons why courts defer to executive and legislative decision making. These can be grouped together into three intersecting principles, namely, the court’s views on the democratically-legitimate role of a court in a constitutional democracy, the court’s views on its appropriate role given its institutional limitations, and the nature of the dispute before a court. Together, these three principles constitute a particular court’s approach to deference. These three principles are discussed separately below.

2.1 Principles of democracy

The first aspect making up a court’s approach to constitutional deference relates to its understanding of the institutional independence or interdependence of the three branches of government, in particular, the role of the courts in a democracy when engaged in the process of judicial review. This is by far the most important normative factor underpinning constitutional deference, and the approach taken by courts on this issue often colours the approach of the courts to the other two principles. This issue has been described by Jowell as one of ‘constitutional competence’, involving ‘a normative assessment of the proper role of institutions in a democracy’.

The democratic legitimacy of judicial review has been the subject of intense academic debate, both historically and currently. It is important to note, at the outset, that this discussion of the debate focuses on whether the institutional practice of judicial review is, in itself, democratic, and not on

2 J Jowell ‘Of vires and vacuums: The constitutional context of judicial review’ in C Forsyth (ed) Judicial review and the constitution (2000) 327 330. Jowell’s argument is set out in relation to administrative judicial review, but is nevertheless applicable to constitutional review.

3 For a good overview of this debate in the US, see B Friedman ‘The birth of an academic obsession: The history of the countermajoritarian difficulty’ (2002) 112 Yale LJ 153.
the efficacy of the protection of rights through a system of judicial review. That is a separate question.4

Neither does this discussion consider the political objections to a rights-based discourse which argues that a rights-discourse impoverishes our conception of society and leads to a preoccupation of the individual with his or her rights, rather than a more group-based approach to rights.5 The argument that judicial review is a preferable or even a necessary means to ensure adequate protection of constitutional rights does not affect the question of the democratic legitimacy of judicial review. Rather, the ‘democracy versus juristocracy debate’ is one that hinges on the countermajoritarian nature of judicial review, since unelected (and therefore democratically unaccountable) judges, in engaging in judicial review, overturn the decisions of the democratically-elected representatives of the majority. This is the ‘paradox’ of judicial review in constitutional democracies: on the one hand, separation of powers requires that courts hold government accountable to the standards set out in the constitution; yet this power given to the courts may be used to thwart the very right to political participation by withdrawing debate from the public arena to the domain of the courts.6

Advocates of judicial review have essentially three arguments open to them in the face of this objection to judicial review. First, they can limit the scope of judicial review to procedural matters in order to remove the democratic objections to substantive judicial review. This is the approach promoted by Ely.7 This argument will not be considered here, and focus will be placed instead on arguments surrounding the democratic legitimacy of substantive constitutional review as these are the arguments relevant to a discussion of deference in constitutional review. The second avenue open to judicial review proponents is to question the assumption that democracy is the sole determinant of legitimacy. Such theorists would accept the undemocratic nature of constitutional review, but argue that it is nonetheless valuable as it affords greater protection for rights. The final option is to challenge a notion of democracy that excludes judicial review, that is, when properly understood, democracy is in fact enhanced by judicial review.8

The debate regarding the democratic legitimacy of judicial review is perhaps best understood by referring to the authors who reflect the extreme positions of the academic spectrum: Ronald Dworkin and Jeremy Waldron. Dworkin relies primarily on the third argument, that is, an expanded notion of democracy, but also puts forward arguments regarding the value of judicial review which could be used to support pragmatic arguments in favour of judicial review. Waldron, on the other hand, asserts the primacy of democracy and attempts to refute Dworkin’s democratic argument for judicial review. It should be noted from the outset, that the Dworkin-Waldron debate is located firmly within the jurisprudence of the United States constitutional democracy, which is, given the lack of direct constitutional authority for judicial review and the extent to which the courts engage in judicial review, not the norm internationally. Indeed, in the United Kingdom and in Canada (and to a much greater extent in South Africa), courts are mandated to undertake judicial review – thus undermining, to some extent, the need and substance of both Dworkin’s and Waldron’s arguments in those jurisdictions. Nevertheless, the debate is informative for discussions on the democratic legitimacy of judicial review.

Waldron argues that there is no necessary connection between the adoption of a ‘rights-based position’ (a term he adopts from Dworkin) to indicate that a concern for fundamental rights lies at the foundation of a particular position\(^9\) and the protection of those rights in a bill of rights with enforcement through judicial review.\(^10\) His central thesis is that judicial review is a negation of the individual right to democratic self-government, which is given effect through electoral representation.\(^11\) Constitutionalisation of rights, he argues, is undermined by its own logic. This is because the entrenchment of a right in a bill of rights and the ‘attitude of mistrust’ that this comprises (since it precludes citizen involvement in the development of rights jurisprudence) is undermined by the underlying premise of rights themselves – that citizens are autonomous and responsible agents.\(^12\)

Waldron emphasises the importance of citizens’ right to participation as stemming from ‘our democratic principles, and from our conviction that self-government and participation in politics by ordinary men and women, on equal terms, is itself a matter of fundamental right’.\(^13\) The entrenchment of constitutional rights amounts to an abrogation of this right to the courts, a move which Waldron finds unacceptable:

---


\(^11\) This argument was first presented in Waldron (n 10 above), and later developed in J Waldron *Law and disagreement* (1999).

\(^12\) Waldron (n 11 above) 249 - 252.

\(^13\) Waldron (n 10 above) 36.
Towards a framework for understanding constitutional deference

[T]his arrogation of judicial authority, this disabling of representative institutions, and above all this quite striking political inequality, should be frowned upon by any right-based theory that stresses the importance of democratic participation on matters of principle by ordinary men and women.14

Instead, Waldron argues that it is only through majority decision making that disagreement can properly be accommodated. Disagreement, for Waldron, is ‘one of the basic circumstances of political life’15 and majoritarian processes are the only way in which to respect these differences. It is precisely because majoritarian processes are based on a rights-based resolution of these differences that it gains its legitimacy and its authority to make law. Majority participation in dispute resolution calls upon the very capacities that rights as such connote, and it evinces a form of respect in the resolution of political disagreement which is continuous with the respect that rights as such evoke.16

On the other end of the spectrum, in Freedom's law,17 Dworkin argues for the democratic legitimacy of judicial review using both a negative and a positive argument. He begins by questioning the majoritarian assumption – that in order for a politically-important decision to count as ‘democratic’ it must be one that the majority would agree to, given enough time and information to make an informed decision; if a decision is not one which the majority would agree to, then it is necessarily undemocratic18 – that judicial review is necessarily undemocratic: ‘Democracy does not insist on judges having the last word, but it does not insist that they must not have it.’19

Dworkin rejects this majoritarian view of democracy in favour of what he calls the ‘constitutional conception of democracy’, where the ‘defining aim of democracy’ is taken to be

that collective decisions be made by political institutions whose structure, composition, and practices treat all members of the community, as individuals, with equal concern and respect.20

Thus, Dworkin offers an account of substantive democracy, where substance takes precedence over form. For Dworkin, it is far more important that rights are enforced correctly, than that rights are interpreted by a majoritarian

14 Waldron (n 10 above) 42.
15 Waldron (n 11 above) 246.
16 Waldron (n 11 above) 252.
18 Dworkin (n 17 above) 16.
19 Dworkin (n 17 above) 7.
20 Dworkin (n 17 above) 17. See also S Fredman ‘Judging democracy: The role of the judiciary under the Human Rights Act 1998’ (2000) 53 Current Legal Problems 99 101-108, who argues that the HRA presents a significant opportunity to enhance participatory democracy in the arena of human rights.
The danger to democracy is in making a wrong decision – not in having the courts make it. Of course, the problem with the argument is the idea that there is one correct answer, or at least, that we can know what that answer is. Reasonable people differ as to how to interpret rights, so it can hardly be an argument in favour of judicial review that the judiciary is better able to find that correct answer.

Dworkin’s positive argument for judicial review focuses on the institutional strengths of the judiciary and the weaknesses of majoritarianism and the legislative process, arguing that despite reasonable disagreements over rights, the institutional structure of judicial review compared with the features of the legislature makes it more likely that judicial review will respect and protect rights and democracy better than the legislature would.

For Dworkin, majoritarian politics ‘encourages compromises that may subordinate important issues of principle’. On the other hand, judicial review of constitutional rights and the widespread public debate which they give rise to, may even, for Dworkin, ‘provide a superior kind of republican deliberation’ to that of majoritarian processes. Dworkin therefore provides a strong counter-argument to the ‘democratic’ argument for majoritarian decision making by developing a more sophisticated understanding of democracy and deliberation.

Hence, in the context of a discussion of judicial activism and judicial restraint, or a ‘policy of deference’, Dworkin tackles what he calls the ‘argument from democracy’ as a justification for a policy of deference, namely, that it is more appropriate for democratically elected legislators to decide issues of moral and political importance. Dworkin challenges this view, pointing out that state legislators (in the United States) are not in fact ‘responsible to the people in the way that democratic theory assumes’. For Dworkin, it is simply not a self-evident argument from the United States constitutional text or practice that courts lack the democratic legitimacy to make policy decisions. After all, the common law itself is derived from the courts.

Dworkin then goes on to deliver a far more fundamental critique, examining the argument on its own terms – that a decision by a democratic institution would be more likely to be sounder and fairer. Dworkin

---

21 Dworkin (n 17 above) 17; Lenta (n 8 above) 10 - 11.
22 Lenta (n 8 above) 17 - 18.
23 Dworkin (n 17 above) 30.
24 Dworkin (n 17 above) 31.
25 The policy of deference must be distinguished from the principle of deference, which applies throughout judicial review. A policy of deference is a particular practice of judicial restraint adopted by a particular court.
26 Dworkin (n 9 above) 140.
27 Dworkin (n 9 above) 141.
28 As above.
points out that the argument that a decision by a majority is always fairer than a decision by a minority ‘ignores the fact that decisions about rights against the majority are not issues that in fairness ought to be left to the majority … [T]o make the majority judge in its own cause seems inconsistent and unjust.’29 Hence, judicial review is part of the process of ensuring that rights are protected and the Constitution is upheld.

From this brief discussion of some of the academic debates surrounding the role of the courts in a democracy, it is clear that there is a range of reasonable and defensible views which a court could adopt. Whatever jurisprudential position courts adopt, it is clear that this position will be crucial in determining a court’s approach to constitutional deference. The Canadian and English courts – which have been explicit about the importance of constitutional competence as a key consideration in prompting deference to executive and legislative decision making – provide a good example.30 In the United Kingdom, with its strongly ingrained tradition of parliamentary sovereignty, courts may be more inclined to adopt a position akin to that of Waldron.31

Canadian courts, on the other hand, appear to be more ready to embrace their ‘democratic’ role in judicial review and see themselves as engaged in the democratic process when undertaking constitutional review. In one of the leading cases on deference in Canada, McLachlin J held:

As with context, however, care must be taken not to extend the notion of deference too far. Deference must not be carried to the point of relieving the government of the burden which the Charter places upon it of demonstrating that the limits it has imposed on guaranteed rights are reasonable and justifiable. Parliament has its role: to choose the appropriate response to social problems within the limiting framework of the Constitution. But the courts also have a role: to determine, objectively and impartially, whether parliament’s choice falls within the limiting framework of the Constitution. The courts are no more permitted to abdicate their responsibility than is parliament. To carry judicial deference to the point of accepting parliament’s view simply on the basis that the problem is serious and the solution difficult, would be to diminish the role of the courts in the constitutional process and to weaken the structure of rights upon which our constitution and our nation is founded.32

29 Dworkin (n 9 above) 142.
30 For a detailed discussion comparing the English and Canadian approaches, see McLean (n * above) 26 - 60.
Canadian courts would consequently be far less likely to defer easily to parliament or the executive in deciding issues of policy than their English counterparts.

The South African Constitution explicitly includes socio-economic rights, and these rights have, from the outset, been accepted as justiciable. Courts have not, therefore accepted arguments that socio-economic rights are not justiciable on the basis of constitutional competence arguments, since courts are not democratically mandated to decide issues of socio-economic policy. Such arguments are, however, nonetheless relevant to the manner in which socio-economic rights matters are decided and the relief which is given.

2.2 Institutional competence

The second consideration underpinning a court’s approach to deference relates to perceived institutional limitations in the various branches of government, and is based on a ‘practical evaluation of the capacity of decision making bodies to make certain decisions’. For Lord Steyn, the ‘relative institutional capacity’ of the courts is the critical factor in deciding whether courts should defer to the other branches of government. This consideration is given further weight in the modern bureaucratic state with its reliance on specialist expertise in almost all areas of the government machine. As a result, some have argued that it is inappropriate for courts to engage in review of complex government policy, or ‘polycentric decision making’, as judges lack the experience, knowledge or resources to make these types of decisions, in particular, to assess what the consequences of their decisions may be and to respond flexibly to unanticipated results of those decisions. For the same reasons, as a general rule, courts will be far more willing to defer to other agencies where the matter is one of fact or policy rather than of law or constitutional interpretation.

It is important to note from the outset that, as with the other factors underpinning constitutional deference, the question of institutional capacity is based on the perceived appropriateness of the courts to make certain types of decisions, rather than an inherent inability to make these decisions. In principle, there are very few, if any, decisions which a court cannot make, if given enough time and information. In commenting on a

34 Jowell (n 2 above) 330.
Towards a framework for understanding constitutional deference

British Court of Appeal judgment dealing with the allocation of limited budgets in the healthcare sector, in which the Court stated that it could not make such a judgment, Lord Justice Dyson noted the following:

I do not think that the court was saying that the court cannot make such a judgment. Clearly, it is not impossible for the court to do so, especially if it is provided with all the material that was available to the decision makers. But it is not the normal function of courts to make such judgments, and they are less-well-equipped than health authorities to make them.36

Thus, the question of institutional competence is not purely one of intellectual capacity and knowledge, but one of appropriateness. In this sense, it is clearly linked to issues of constitutional competence. Another issue affecting the institutional competence of courts to adjudicate certain matters is the nature of the evidence placed before the court. Where evidence relates to complex social science matters, courts will be more deferent to the decision making of the other branches of government.37 There is also, of course, a limit to the quantity of evidence that can be placed before the court and which judges can be expected to assimilate.38

Lon Fuller’s famous article on polycentricity is often cited as authority for the view that courts are not appropriate forums to decide sufficiently polycentric matters.39 Since decisions involving social policy or socio-economic rights are normally thought to involve polycentric issues and complex issues of policy, courts are often reluctant to adjudicate on such matters and will usually accord the state a high level of deference in such adjudication.40 In particular, where there are competing socio-economic theories, courts will be extremely reluctant to make policy choices. Given the centrality of Fuller’s article on this matter, it is worth discussing it in some detail.

The notion of polycentricity was first introduced and discussed by Fuller in his posthumously published article of 1978, ‘The forms and limits of adjudication’. This article remains the foremost exposition on polycentricity, cited repeatedly in the secondary literature and case law as authority for the proposition that polycentric decisions are not justiciable, with little further discussion. Fuller’s article explores the use of adjudication as a form of social ordering, contrasting adjudication with two

37 See the minority decisions in MacDonald (n 32 above) which discuss this consideration in detail.
40 Decisions in Canada and the United Kingdom illustrate this approach well: See the minority decisions in MacDonald (n 32 above), Chaoulli v Quebec (AG) (2005) 1 SCR 791; and the English decisions in Kebilene (n 31 above); Roth (n 31 above) paras 144 184.
other forms of social ordering, namely, contract and voting. He argued that various types of decisions are ideally suited to one of these types of social ordering, and in so doing, sought to draw out the limits of adjudicative decision making. In particular, Fuller argued that a polycentric matter is one that is not well-suited to adjudicative resolution, since a polycentric matter is one in which there is a matrix of interrelated issues, where the effect of altering one of the factors would have unpredictable consequences on the rest of the issues in the matrix. He used the by now well-known illustration of the spider’s web, where the plucking of one strand on the web would result in a complex set of changes throughout the web:

We may visualise this kind of situation by thinking of a spider web. A pull on one strand will distribute tensions after a complicated pattern throughout the web as a whole. Doubling the original pull will, in all likelihood, not simply double each of the resulting tensions but will rather create a different complicated pattern of tensions. This would certainly occur, for example, if the doubled pull caused one or more of the weaker stands to snap. This is a 'polycentric' situation because it is 'many centered' – each crossing of strands is a distinct center for distributing tensions.\(^{41}\)

Fuller derived his idea of polycentricity from the work of Polanyi, who developed the concept of polycentricity to argue that there are inherent informational constraints on any person making a central decision on a complex set of facts. For this reason, polycentric central state decisions regarding budgetary allocation are inappropriate, and should be left to market forces and 'spontaneous mutual adjustment'.\(^{42}\)

Individuals within the system should ‘evaluate by their independent mutual adjustments the polycentric task of optimum allocation of resources and distribution of products’ to resolve complex polycentric tasks, rather than attempt to solve the task centrally.\(^{43}\) Fuller used Polanyi’s concept of polycentricity to argue that when a matter is significantly polycentric, it becomes problematic for the issue to be resolved through a centralised decision-making process. His argument was directed primarily at curbing the use, by legislatures, of administrative agencies to resolve complex matters with significant polycentric consequences.\(^{44}\) His objections to this type of decision making are, however, relevant also to the judicial resolution of disputes with significant polycentric implications.

It is important to note that polycentricity, or non-polycentricity, are not absolute categories. Rather, polycentricity is a matter of degree, and most decisions before a court have some elements of polycentricity. The greater the degree of polycentricity, the more difficult the decision will be

---

41 Fuller (n 39 above) 395.
42 M Polanyi The logic of liberty: Reflections and rejoinders (1951) 170 - 184.
43 Polanyi (n 42 above) 179.
and the less appropriate the matter becomes for judicial resolution. As Fuller recognised, concealed polycentric elements are probably present in almost all problems resolved by adjudication ... It is a question of knowing when the polycentric elements have become so significant and predominant that the proper limits of adjudication have been reached.\(^{45}\)

This, in turn, will mean that courts tend to be reluctant to become engaged in matters that they regard as sufficiently polycentric, and will find them to be non-justiciable. As Fuller notes, however, a high degree of polycentricity cannot be a bar to judicial resolution in itself, and there may be instances where it is preferable for a court to engage in the resolution of a polycentric matter where it cannot be adequately resolved through other means.\(^{46}\)

In this discussion of Fuller’s work, the focus is on three important issues implicated by the judicial resolution of polycentric disputes – issues which are relevant for the determination of justiciability and the appropriate level of deference to be applied by a court in a particular matter. The first issue is the lack of evidence before, or lack of expertise within, a court to enable it to predict accurately the consequences of its decisions.\(^{47}\) Moreover, if parties who may be adversely affected are not before the court – even if they could be identified – it may not be practical for a judge to attempt to take all their interests into account. When a court is confronted with having to decide a complex, polycentric matter it should therefore, in Allison’s words, avoid choosing to decide the matter in a way that results in a decision which ‘necessitates an appreciation of complex repercussions’; nor should it develop the law where a similar appreciation of the repercussions is required.\(^{48}\) It will not, however, always be clear to the court whether a matter before it is one which involves a complex, polycentric determination, precisely because the evidence before it is limited to that which the parties place before it. For Allison, Fuller’s concept of polycentricity therefore becomes too vague to use as a principle of adjudication.\(^{49}\) Allison also criticises Fuller for failing to explain what the alternative to judicial resolution of polycentric matters is. The other social institutions identified by Fuller (the legislature, mediation, contract, and managerial decision making) are similarly not ideally suited to deciding polycentric matters – although Fuller does argue that contract and managerial direction are superior. Neither does Fuller adequately explain when it would be preferable, as a matter of principle, for courts nevertheless to engage in polycentric decision-making.\(^{50}\) Thus, while Fuller’s concept of

\(^{45}\) Fuller (n 39 above) 398.

\(^{46}\) Fuller (n 39 above) 405 - 406. Allison notes that in his correspondence with Frank Newman, Fuller accepted that in certain situations, decisions such as the desegregation decisions of the 1950s were necessary, but that they place a ‘serious moral drain on the integrity of adjudication’: Allison (n 44 above) 374.

\(^{47}\) Fuller (n 39 above) 401.

\(^{48}\) Allison (n 44 above) 370 - 371.

\(^{49}\) Allison (n 44 above) 372 - 373.

\(^{50}\) Fuller (n 39 above) 398 - 400; Allison (n 44 above) 373 - 374.
Polycentricity has important consequences for judicial adjudication of polycentric matters, it is difficult to see how it can be used as a principled basis for deciding which matters are, or should be, justiciable.

The second important issue raised by Fuller is the role of the judge. Fuller's conception of adjudication coloured his understanding of the role of the judge. For Fuller, a judge is more like an umpire who makes decisions on the basis of the evidence presented by the parties.51

Adjudication is, then, a device which gives formal and institutional expression to the influence of reasoned argument in human affairs. As such it assumes a burden of rationality not borne by any other form of social ordering. A decision which is the product of reasoned argument must be prepared to meet the test of reason. We demand of an adjudicative decision a kind of rationality we do not expect of the results of contract or of voting. This higher responsibility toward rationality is at once the strength and the weakness of adjudication as a form of social ordering.52

For Fuller, the only way to ensure that the judiciary will not pre-judge a matter or be biased is through the adversarial procedure. As a result, he failed to consider the potential of a more inquisitorial approach, as used in many continental judicial procedures, for example, as an alternative to his strictly adversarial approach.53 It is arguable that some of the deficiencies in the adjudication process identified by Fuller could be ameliorated if the judge in a polycentric dispute adopted a more inquisitorial role.54 In such a role, a judge would be able to help identify polycentric issues and protect the interests of potentially affected parties which are not before the court.55 The Indian courts provide a good example of how this could be done: These courts have instituted mechanisms to ensure that adequate evidence is placed before them through the appointment of 'socio-legal fact-finding' commissions.56 Similarly, courts can appoint an amicus curiae to represent the interests of those not directly before the court, or request a state body, or even a non-governmental organisation, to make representations to the courts.

51 Fuller (n 39 above) 365 - 367.
52 Fuller (n 39 above) 366 - 367.
53 Allison (n 44 above) 377, 380. It was this 'Anglo-American bias', recognised by Fuller, which was one of the reasons Fuller was not prepared to publish his article: Allison (n 44 above) 377.
54 In South Africa, eg, courts are not unfamiliar with an inquisitorial approach, as there is a statutory obligation to adopt an inquisitorial approach in criminal law proceedings, where this is necessary to obtain a just decision: Rex v Hepworth [1928] AD 265.
The third important argument, which derives from Fuller’s work, is that it is very difficult for courts to choose between two equally valid policy choices. 57

While, it must be acknowledged that this statement is, on the whole, accurate, a couple of observations arise from this point. First, it may be just as difficult for a legislature or executive to make these choices. While these institutions potentially have greater access to data, they are not able to make perfect decisions based on perfect knowledge either. Indeed, the very idea of polycentricity was developed by Polanyi to argue against centralised state planning and to argue in favour of market-generated planning. The crucial difference is that the legislature and executive are democratically mandated to make these decisions – perfect or not – but that is a separate objection to judicial consideration of polycentric issues. The second point is that involving the judiciary in judicial review of social policy does not necessarily mean that the courts have to make policy choices to the exclusion of the executive and legislature, and can engage in a dialogue with the other branches of government. As the South African jurisprudence, for instance, demonstrates, courts are able to assess state policy for reasonableness, and in this way, engage in a dialogue over the development of policy in line with constitutional values. 58

In short, Fuller raises a number of important issues regarding the adjudication of polycentric decisions which are relevant to a discussion of constitutional deference. Nevertheless, it should be remembered that Fuller’s work is limited in its scope and fails to deal with many issues. According to Allison, Fuller himself never regarded his article as sufficiently complete to be ready for publication. He recognised that he had failed adequately to take into consideration other forms of adjudication, developments in public law as a result of the civil-rights litigation, and that his description of adjudication was problematic. 59 Indeed, in his later work dealing with judicial review, he did not deal with polycentricity at all. 60 Fuller certainly raised important issues, but his analysis can by no means be considered the last word on the matter. Polycentricity is therefore not a bar to justiciability, but merely one consideration to be taken into account by the judiciary in deciding whether a matter should be justiciable or what the appropriate level of constitutional deference should be.

57 Pieterse (n 55 above) 393. See the discussion of Makwanyane (n 78 above).
59 Allison (n 44 above) 377 - 378.
The institutional competence of the courts is the consideration on which the South African courts set most store in the adjudication of socio-economic rights. As discussed in more detail below, this contrasts with the approach of the courts to the adjudication of civil and political rights, where the Constitutional Court has held that deference to the 'practical difficulties' faced by the administration, does not mean that decision makers should not be held to account for infringement of constitutional rights: 'It is the remedy that must adapt itself to the right, not the right to the remedy.'

In socio-economic rights adjudication, by contrast, it appears to be that the right is shaped by the remedy.

2.3 The nature of the subject matter

The third consideration relates to the nature of the subject matter under review. There are a number of ways in which this can affect the level of deference applied by a court. First, where the action is one characterised by greater political discretion, courts will tend to be more deferential to the decision-making of the executive or legislature. One of the clearest decisions demonstrating this consideration is found in the British decision of *A v Secretary of State for the Home Department*, where Lord Bingham expressly stated that one of the reasons for affording deference to the executive in its conclusion that the United Kingdom is facing a 'threat to the life of the nation' is because this is a 'pre-eminently political judgment':

The more purely political (in a broad or narrow sense) a question is, the more appropriate it will be for political resolution and the less likely it is to be an appropriate matter for judicial decision. The smaller, therefore, will be the potential role of the court. It is the function of political and not judicial bodies to resolve political questions. Conversely, the greater the legal content of any issue, the greater the potential role of the court, because under our Constitution and subject to the sovereign power of parliament it is the function of the courts and not of political bodies to resolve legal questions.

This simple distinction between 'legal' and 'political' questions, however, threatens to unravel, particularly where two competing rights or interests are at play. The example above, for instance, begs the question: Why is the interpretation of what constitutes a 'threat to the life of the nation' purely a political question where it also constitutes a statutory pre-condition for legitimate derogations from the HRA? Surely this could be characterised equally as a legal question? The decision by a court to label a matter 'political', therefore, will often be used as a mask for a prior decision to afford the decision-maker a high level of deference, based on one of the other two factors, or even on a non-principled factor, such as an unwillingness to

---

61 Bel Porto School Governing Body v Premier, Western Cape 2002 3 SA 265 (CC) para 186.
62 n 31 above, para 29.
become involved because of the social or political consequences of doing so.63

Nevertheless, in principle, the political nature of the decision may legitimately form the ground for a deferent stance by the courts.

A second instance where courts are more likely to be deferential to the decisions of the executive or legislature is where the constitution or right in question permits a wide range of legitimate responses, or where the right has to be balanced against another right. Young calls this ‘substantive legislator deference’.64 Examples of such situations can be found in the case law of Canada and the United Kingdom. In Canada, in both Edwards Books and Irwin Toy,65 for instance, the key factor underpinning the Supreme Court’s approach to deference was the fact that the legislative decision under review was aimed at reaching a compromise between the rights of two competing groups. Similarly, in Kebilene and Roth,66 the House of Lords noted that where rights are unqualified, courts are well placed to determine the content of the right and the legality of any infringement to that right; but where the Convention requires a balancing of rights, greater deference is due.

A variation on this theme has been enunciated by the South African Constitutional Court in Ferreira v Levin,67 where Ackermann J, in a discussion of the broad, residual nature of the right to freedom and security of the person in section 11 of the interim Constitution, acknowledged that the German Federal Constitutional Court was more deferent to the German legislature in areas protected by the general right of ‘freedom of action’ in contrast to other rights which are more expressly protected. This approach, the Court noted, is analogous to the United States ‘heightened scrutiny’ of ‘fundamental rights’. In short, Ackermann J found that where a right is expressly and narrowly protected in the Constitution, the Court would be less deferent to the legislature than would be the case where the interest was protected generally or through a residual right. Hence, where a right is given express and precise protection, courts will be less deferent in protecting that right.

Third, when fundamental rights, highly prized in a particular society, are at issue, a court is less likely to defer to choices made by the agency in question. This approach often results in courts drawing lines between what

66 Kebilene (n 31 above) 380; Roth (n 31 above) 84, per Laws LJ in dissent.
67 Ferreira v Levin NO; Vryenhoek v Powell NO 1996 1 SA 984 (CC).
they will protect as ‘fundamental human rights’ and those which are non-justiciable or which require a greater degree of deference. This approach of differing levels of review is notably expounded by Justice Stone in the United States Supreme Court in *United States v Carolene Products Co* in the fourth footnote. In that footnote, Justice Stone explained that the new deferential method expounded by the Court in that case did not mean that the same level of deference was necessarily to be applied in all constitutional litigation and that legislation aimed at ‘particular religious, or national, or racial minorities’, for example, may call for a ‘more searching judicial enquiry’.68 This approach has been adopted in the United Kingdom, for instance, where the House of Lords, in *R v Carson*, affirmed that in equality matters ‘severe scrutiny’ is appropriate.69 Similarly, in South Africa, the Constitutional Court has stated no institutional deference is necessary or desirable where a court is to determine whether the right to equality has been infringed.70

Socio-economic rights, by contrast, are a good example of a category of rights where many judges would accept that a highly deferent approach should be adopted – that is, if they are to be considered justiciable at all. In the same vein, decisions with resource-allocation implications are another category of cases where courts will generally show a great degree of deference to the decision making of the executive or legislature.71 As Fredman points out, however, just as the distinction between ‘legal’ and ‘political’ decisions is dubious, the category of ‘social or economic’ decisions, or those with resource implications, is difficult to sustain.72

In addition to these three contextual factors, others can be postulated. For instance, it is arguable that the subject matter should affect the level of review imposed where the agency has an interest in the outcome and may be perceived as biased in the decision-making process. In such cases, it is important for the court to be seen as an independent arbiter. A good example is where political rights are involved, such as the right to vote. Where a dispute arises around voting regulations or practices, a court should be quick to adjudicate the matter and apply a high level of scrutiny to the actions of the agency. This justification arises out of a more pragmatic conception of separation of powers.

---

70 *MEC for Education, KwaZulu-Natal v Pillay* 2008 1 SA 474 (CC) para 81.
72 Fredman (n 63 above) 58 - 59.
3 Constitutional deference in the adjudication of socio-economic rights

The second part of this article seeks to apply the framework set out above to an analysis of the adjudication of socio-economic rights. Before doing so, a brief overview of the development of the jurisprudence on socio-economic rights is provided.

3.1 A brief overview of the rhetoric of deference in South Africa

The South African judiciary, under apartheid, has been roundly criticised for its ‘executive-mindedness’, its failure to uphold basic rights and its overly deferential approach to executive political will.73 Most Appellate Division judges, steeped in the culture of parliamentary sovereignty, characterised their task as being solely to interpret and apply the intention of the legislature, with little or no regard for individual liberties, thereby allowing ‘a judge in covert sympathy with a legislative programme to give full effect to his predispositions without having to accept public responsibility for doing so’.74 The reality for many in the judiciary was at best a tacit complicity in apartheid policies and, at worst, a thinly-veiled support for that system.75 For these reasons, some commentators have questioned the wisdom of contemporary calls for a theory of deference in post-apartheid adjudication.76

The term ‘deference’ is used, nonetheless, in this article, in spite of its association with apartheid decision making. The main reasons in favour of doing this are that it helps to ensure that South African jurisprudence shares the terminology of other commonwealth jurisdictions, and that it seeks to ‘reclaim’ the term from its unduly negative associations. The theory of constitutional deference advocated in this article is markedly different to the pro-executive, and sometimes subservient, attitude adopted by the pre-democratic judiciary. Rather, it is a framework in which to interpret and critique all judicial decision-making.


74 Cameron (n 73 above) 60.

75 Davis & Corder (n 73 above) 295 - 302.

Pre-democratic uses of the term ‘deference’ in South African judgments are mostly to be found where courts note the need to be appropriately deferent to lower courts, or tribunals with regard to factual and credibility findings of those decision-making bodies. This is a form of deference to the institutional competence of the lower courts and tribunals as they are best placed to make certain findings, and continues to this day. Deference to the ‘fact-finder’ is excluded from consideration in this article. Post-apartheid courts, by contrast, have begun, tentatively, to expand their use of the notion of deference to the more sophisticated sense used in comparative jurisdictions (notably those of Canada and the United Kingdom). The remainder of this section is devoted to a discussion of the most important of these decisions, illustrating the development of this jurisprudence.

The first significant judgment to consider the notion of deference was \textit{S v Makwanyane}, which abolished the death penalty.\footnote{See, eg, \textit{Commissioner for Inland Revenue v Louw} 1983 3 SA 551 (A) 569; \textit{Ndolvu v AA Mutual Insurance Association Ltd} 1991 3 SA 655 (E) 659. Prior judgments used the word as a synonym for ‘respect’, and generally in relation to other judges or academics.} Chaskalson P, with reference to the Canadian decision of \textit{Tetreault-Gadoury v Canada},\footnote{\textit{S v Makwanyane} 1995 3 SA 391 (CC).} noted that, where choices have been made between ‘differing reasonable policy options’, courts must give the legislature a measure of deference in that choice. He cautions, however, that this deference does not afford the legislature an ‘unrestricted licence’ to infringe constitutional rights, and the state must still show a reasonable basis for the limitation of the right.\footnote{\textit{Tetreault-Gadoury v Canada} (Employment and Immigration Commission) (1991) 4 CRR (2d).} Similarly, in \textit{Ferreira v Levin} 81 (discussed above), Ackermann J held that courts should be less deferent where rights are expressly and narrowly protected. Hence the specificity with which the right is framed, will be influential in determining the degree of deference afforded by the courts to the legislature or executive. Notably, both judgments concerned civil and political rights.

In a discussion of the appropriate remedy to cure a constitutional defect in the \textit{National Coalition} 82 decision, the Constitutional Court again used the language of deference, finding that the deference owed to the legislature in deciding what constitutes appropriate relief will depend on the individual circumstances of each case. The Court held:

\begin{quote}
It is not possible to formulate in general terms what such deference must embrace, for this depends on the facts and circumstances of each case. In essence, however, it involves restraint by the courts in not trespassing onto that part of the legislative field which has been reserved by the Constitution, and for good reason, to the legislature. Whether, and to what extent, a court may interfere with the language
\end{quote}
of a statute will depend ultimately on the correct construction to be placed on the Constitution as applied to the legislation and facts involved in each case.83

This passage illustrates the Court’s awareness of the context-sensitive nature of the choice of remedies, and the role that deference to the legislature plays in that assessment. It also demonstrates a concern for institutional competence in its reference to the ‘field’ which is reserved for the legislature. This passage was later quoted by the same court in *UDM v President of the Republic of South Africa*84 in a discussion of what would constitute ‘appropriate relief’ in any given case.

The Constitutional Court again emphasised its institutional limitations in the decision of *Bel Porto School Governing Body*,85 but noted that an appreciation of these limitations should not undermine the court’s role in interpreting and protecting rights. The Court found that, while courts should, as a general principle, be deferent to the ‘practical difficulties’ faced by the administration, this does not mean that decision makers should not be held to account for infringement of constitutional rights: ‘It is the remedy that must adapt itself to the right, not the right to the remedy’.86

Similarly, in a discussion of the role of the judiciary in the review of administrative action which involved polycentric decision making, Cameron JA, in the Supreme Court of Appeal decision of *Logbro Properties*,87 developed the Constitutional Court’s jurisprudence of institutional competence, holding that a ‘measure of judicial deference’ is appropriate to the administration. The Court went on, however, to link institutional competence concerns with democratic competence issues. Deference is important, Cameron JA held, to maintain the distinction between review and appeal, to ensure that the judiciary ‘appreciate[s] the legitimate and constitutionally-ordained province of administrative agencies’, recognising their expertise in such matters, and to ensure that they are ‘sensitive in general to the interests legitimately pursued by administrative bodies and the practical and financial constraints under which they operate’.88

*Logbro Properties* was then developed by the same court in its later decision of *Phambili Fisheries*,89 where the Court held that deference in the judicial review of government economic policies was appropriate for the

---

83 *National Coalition* (n 82 above) para 66.
84 *United Democratic Movement v President of the Republic of South Africa (African Christian Democratic Party Intervening; Institute for Democracy in South Africa as Amicus Curiae)* (no 2) 2003 1 SA 495 (CC) para 115.
85 *Bel Porto* (n 61 above).
86 *Bel Porto* (n 61 above) para 186.
89 *Minister of Environmental Affairs and Tourism v Phambili Fisheries (Pty) Ltd; Minister of Environmental Affairs and Tourism v Bato Star Fishing (Pty) Ltd* 2003 6 SA 407 (SCA).
same institutional and democratic competence reasons. Schutz JA went on to state that ‘[j]udicial deference does not imply judicial timidity or an unreadiness to perform the judicial function. It simply manifests the recognition that the law itself places certain administrative actions in the hands of the executive, not the judiciary’. This is particularly important, he held, where the subject matter under review is ‘very technical or of a kind in which a Court has no particular proficiency’. Schutz’s emphasis on institutional competence was picked up in two High Court decisions. In Foodcorp, the Cape High Court, following both Logbro Properties and Phambili Fisheries, acknowledged the importance of ‘due judicial deference’ to ‘policy-laden and polycentric’ administrative action which ‘entails a degree of specialist knowledge and expertise that very few, if any, judges may be expected to have’. In South African Jewish Board of Deputies, the Witwatersrand Local Division, also referring to the judgment of Schutz JA in Phambili Fisheries, held that where a judicial review matter was not one which was ‘very technical or of a kind in which a court has no particular proficiency’, judicial deference was not appropriate at all.

This rather narrow approach to institutional competence in the High Courts was later expanded by the Constitutional Court in Bato Star (the appeal against the Supreme Court of Appeal decision in Phambili Fisheries to the Constitutional Court). In that decision, O’Regan J repeated the passages cited by Schutz JA on the deference to be adopted in the judicial review of administrative agencies. She went on to say, with reference to the ProLife Alliance decision, that ‘the need for courts to treat decision makers with appropriate deference or respect flows not from judicial courtesy or etiquette but from the fundamental constitutional principle of the separation of powers itself’. This reasoning reveals an appreciation for the limits of constitutional competence, and the decision of Bato Star has since become one of the leading decisions on the topic of deference, and has been cited in a number of subsequent judgments.

90 Phambili (n 89 above) para 47.
91 Phambili (n 89 above) para 50.
92 Phambili (n 89 above) para 53.
93 Foodcorp (Pty) Ltd v Deputy Director-General, Department of Environmental Affairs and Tourism, Branch Marine and Coastal Management 2004 5 SA 91 (C) para 68.
94 South African Jewish Board of Deputies v Sutherland NO 2004 4 SA 368 (W).
95 Sutherland (n 94 above) para 38, citing Phambili Fisheries (n 89 above) para 53.
96 Bato Star Fishing (Pty) Ltd v Minister of Environmental Affairs 2004 4 SA 490 (CC).
97 ProLife Alliance (n 31 above) paras 75 - 76.
98 Bato Star Fishing (n 96 above) para 46.
99 See, eg, Associated Institutions Pension Fund v Van Zyl 2005 2 SA 302 (SCA) para 39, where the Court held that deference was appropriate, in the sense used in Bato Star, to the methodology of an expert actuary; Foodcorp (Pty) Ltd v Deputy Director-General, Department of Environmental Affairs and Tourism: Branch Marine and Coastal Management 2006 2 SA 191 (SCA) para 12; Foodcorp (Pty) Ltd v Deputy Director-General, Department of Environmental Affairs and Tourism: Branch Marine and Coastal Management 2006 2 SA 199 (C) 210;
An important refinement by the Constitutional Court is to be found in the recent decision of Pillay, concerning the right of a school girl to wear a nose stud in school as part of her right not to be discriminated against on the basis of religious and cultural practices. The school authorities argued that the Court should show a measure of deference to school ‘governing bodies that are statutorily required to run schools and have the necessary expertise to do so’. In making this argument, they invoked the doctrine of margin of appreciation used by the European Court of Human Rights. The Constitutional Court rightly rejected this argument, pointing out that it had previously decided that the doctrine ‘is not a useful guide when deciding either whether a right has been limited or whether such a limitation is justified’. The Court held that, while judicial deference is appropriate in the review of administrative action where the decision maker is especially well qualified to decide a particular matter, no institutional deference is necessary or desirable where a court is to determine whether the right to equality has been infringed. Specifically, the Court held:

The question before this Court, however, is whether the fundamental right to equality has been violated, which in turn requires the Court to determine what obligations the school bears to accommodate diversity reasonably. Those are questions that courts are best qualified and constitutionally mandated to answer. This Court cannot abdicate its duty by deferring to the school’s view on the requirements of fairness. That approach is obviously incorrect for the further reason that it is for the school to show that the discrimination was fair. A court cannot defer to the view of a party concerning a contention that that same party is bound to prove.

The Constitutional Court’s approach to civil and political rights (exemplified in Pillay) stands in contrast to its record in the adjudication of socio-economic rights. In Minister of Health v TAC, the Constitutional Court for the first time expressly considered the question of how deference and the doctrine of separation of powers relates to its adjudication of socio-economic rights. The Court did so by raising two preliminary issues. The first concerned the deference the Court was to show the executive regarding policy formulation and the second related to the remedy which the Court should provide. In the words of the Court:

---

Intertrade Two (Pty) Ltd v MEC for Roads and Public Works, Eastern Cape 2007 6 SA 442 (C) para 46; Tantoush v Refugee Appeal Board 2008 1 SA 232 (T) para 108.

Pillay (n 70 above) para 80.

Pillay (n 70 above) para 80 (footnotes omitted). The Court referred to the National Coalition (n 83 above) and Makwanyane (n 78 above) decisions as authority for this. See R v Secretary of State for the Home Department, ex parte Daly [2001] 2 WLR 1622 paras 26-27 for a discussion by the British courts as to why the doctrine of margin of appreciation adopted by the European Court of Human Rights is not appropriately applied within their domestic context.

Minister of Health v Treatment Action Campaign (no 2) 2002 5 SA 721 (CC).
These considerations are relevant to the manner in which a court should exercise the powers vested in it under the Constitution. This case is notable for it is the first time the Court expressly considered the question of how deference, or the doctrine of separation of powers, relates to the approach which it should adopt in the adjudication of socio-economic rights rather than the broader, but related, question of whether it should adjudicate them.

In a discussion of why the Court should not adopt a minimum core interpretation of sections 26(1) and 27(1), it found that:

[c]ourts are not institutionally equipped to make the wide-ranging factual and political enquiries necessary for determining what the minimum-core standards ... should be, nor for deciding how public revenues should most effectively be spent.

... Courts are ill-suited to adjudicate upon issues where court orders could have multiple social and economic consequences for the community. The Constitution contemplates rather a restrained and focussed role for the courts, namely, to require the state to take measures to meet its constitutional obligations and to subject the reasonableness of these measures to evaluation ... In this way the judicial, legislative and executive functions achieve appropriate constitutional balance.

On the second issue of the appropriate remedy to be granted, the Court addressed a similar argument made by the state, also under the separation of powers doctrine, that the only appropriate remedy is for the Court to issue a 'declaration of rights', as it is the executive's sole prerogative to make policy.

The Court responded as follows:

This Court has made it clear on more than one occasion that although there are no bright lines that separate the roles of the legislature, the executive and the courts from one another, there are certain matters that are pre-eminently within the domain of one or other of the arms of government and not the others. All arms of government should be sensitive to and respect this separation. This does not mean, however, that courts cannot or should not make orders that have an impact on policy.

The Court was therefore clear that it would not adopt a rigid adherence to a functionalist separation of powers and would, where necessary, make declaratory or mandatory orders that impinge on the traditional roles of the other branches of government, when this is 'mandated by the Constitution itself'.

105 TAC (n 104 above) para 22.
106 TAC (n 104 above) paras 37 - 38.
107 TAC (n 104 above) paras 96 - 97.
108 TAC (n 104 above) para 98 (footnote omitted).
109 TAC (n 104 above) para 99.
A number of other judgments use the language of deference in reference to the doctrine of comity. In *Kaunda*,\(^{110}\) for instance, Chaskalson CJ held that the South African government is entitled to a measure of deference in deciding when, and if, it would seek assurances from another country’s government that they would not impose the death penalty on South African nationals.\(^{111}\) Deference is also employed to discuss the relationship between the courts and the legislature in the development of the common law. In the Supreme Court of Appeal decision in *Fourie*,\(^ {112}\) Cameron JA held that, because the order entailed developing the common law and did not involve any statutory provisions, it was not necessary for the Court to defer to ‘the particular functions and responsibilities of the legislature’.\(^ {113}\) This reasoning may be distinguished from that in *RPM Bricks*,\(^ {114}\) where Ponnan JA held that the common law should not be developed by the courts to allow estoppel to be raised against government where the effect of this would be to render a statutorily-prescribed *ultra vires* act valid. To do so would not show the deference due to legislative authority in prescribing the procedures to be followed before certain powers could be exercised.\(^ {115}\)

In other judgments, courts have recognised the role of the legislature in the South African democracy, and held that deference is due to the legislature by the other branches of government, and that a court should only interfere where it is ‘absolutely necessary to avoid likely irreparable harm and then only in the least intrusive manner possible with due regard to the interests of others who might be affected by the impugned legislation’.\(^ {116}\)

Hence, the post-1994 South African courts have quickly developed the beginnings of a jurisprudence on the doctrine of deference. Yet, this development is uneven: While the application of the doctrine is principled in the interpretation and enforcement of civil and political rights (such as in the *Bel Porto* and *Pillay* decisions), this has yet to be mirrored in its application to socio-economic rights. It is this latter point which is developed in the remainder of this article.

\(^{110}\) *Kaunda v President of the Republic of South Africa* 2005 4 SA 235 (CC).

\(^{111}\) *Kaunda* (n 110 above) para 102.

\(^{112}\) *Fourie v Minister of Home Affairs* 2005 3 SA 429 (SCA).

\(^{113}\) *Fourie* (n 112 above) para 46. This reasoning in the SCA judgment was repeated by the Constitutional Court in *Minister v Home Affairs v Fourie (Doctors for Life International, Amicus Curiae), Lesbian and Gay Equality Project v Minister of Home Affairs* 2006 1 SA 524 (CC) para 21.

\(^{114}\) *City of Tshwane Metropolitan Municipality v RPM Bricks (Pty) Ltd* 2008 3 SA 1 (SCA).

\(^{115}\) *City of Tshwane* (n 114 above) para 24.

\(^{116}\) *President of the Republic of South Africa v United Democratic Movement (African Christian Democratic Party Intervening; Institute for Democracy in South Africa as Amicus Curiae)* 2003 1 SA 472 (CC) para 31. This passage was quoted with approval in *Minister of Home Affairs v Eisenberg and Associates: In re Eisenberg and Associates v Minister of Home Affairs* 2003 5 SA 281 (CC) para 69.
3.2 Socio-economic rights adjudication

The Constitutional Court has made it clear that the socio-economic rights entrenched in the Constitution are justiciable. Despite this, the Court has not dealt with the adjudication of socio-economic rights in the same way that it has treated civil and political rights. This is evidenced primarily through the adoption of a lower standard of review, and deferential remedies. The reasons for this deferential approach appear to be that the Court, despite acknowledging the equal status of socio-economic rights, retains a number of unarticulated and unexamined reservations to the adjudication of these rights. The notion of constitutional deference can be used to examine and critique these reservations. This final section ends with three general observations on the Constitutional Court’s approach to interpretation and enforcement of socio-economic rights based on the framework set out in this article.

First, it is clear that the test adopted by the Constitutional Court to review socio-economic rights is the reasonableness test – a standard of review lower than the standard of proportionality used for the review of civil and political rights. This test is used as it allows the courts to support a restrictive approach to the interpretation and review of socio-economic rights. Following on from this point, the reasonableness test also allows the courts to avoid engagement with a normative interpretation of the scope of the right, demonstrated in the Court’s refusal to entertain an interpretation of the right which includes a minimum core, in Grootboom and TAC.117 Hence, a feature of the Court’s approach is its restrictive interpretation of socio-economic rights and rejection of a more expansive reading of the right found in international law and domestic legislation.

Second, because of the Court’s preference for the reasonableness test used to interpret sections 26 and 27 of the Constitution, it has shown a tendency to attempt to adjudicate all socio-economic rights matters brought before it using this test derived from those sections. Hence, in Soobramoney, the Court refused to consider an expansive reading of the unqualified right to life where the applicant’s claim could be decided under the more restrictive right of access to healthcare in section 27(1).119 This allowed the Court to be more deferent to the hospital policy decision makers, as all they had to demonstrate was that their policy was reasonable, rather than that their policy was a justified limitation of the right to life under section 36 (the limitations clause) of the Constitution. Similarly, in Grootboom (and to a lesser extent in TAC),121 the Court decided claims based on section 28(1)(c) (unqualified children’s socio-economic rights) under sections 26

---

118 TAC (n 104 above) paras 37 - 38.
120 Grootboom (n 117 above) para 71.
121 TAC (n 104 above) paras 74 - 79.
and 27 respectively, thereby watering down children's rights to socio-economic goods. In *TAC*, the Court decided the constitutionality of a violation of a negative aspect of the right to healthcare using the reasonableness test, rather than a more exacting level of scrutiny.\(^{122}\) And finally, in *Khosa*, the Court used the reasonableness test to interpret the term 'everyone' in the manner of a limitations analysis.\(^{123}\)

The final observation relates to the remedies used. It is in the arena of remedies that the court has, arguably, been most deferent. This has resulted in the state often failing to implement an order effectively, or timeously. Most discussion of deference in the case law and secondary literature, both in South Africa and comparatively, considers deference primarily in relation to the interpretation, and limitations or proportionality analysis, of rights determinations. Deference also operates, however, in the remedies that are provided, and it is important to tease out the differences in approach in these two aspects of the adjudication process. Moreover, the reasons for adopting a particular approach to constitutional deference in interpretation and enforcement respectively may differ.

There are two main ways in which the approach to interpretation and enforcement can relate: Either they can reinforce the other, or they can counteract the effect of the other. In the first, the approach to constitutional deference in interpreting the right or statute in question is mirrored by the approach to constitutional deference in the remedy. Thus, for instance, if a court adopts a highly deferential approach in interpreting a right restrictively, and in assessing the evidence before it, it may similarly be deferential in granting an unobtrusive remedy, such as a declaration of unconstitutionality.

The second, and more interesting, relationship is that where the approach to the one is used to counteract the other. For example, where a court adopts a low level of deference to the interpretation of a right and level of scrutiny, and then finds that the state has infringed the right, it may, for reasons of constitutional deference, adopt a highly deferential remedy, or *vice versa*. One could also speculate that where a court, or similar institution, is bound by its constitution to adopt a deferential remedy, or where its findings are not binding, it may be less deferential in its interpretive enquiry and scrutiny of evidence. An example can be found in the United Kingdom, where British courts may make orders of incompatibility only where they find an infringement of the Human Rights Act (although, in practice, these are always acted upon by the government). The South African courts seem to have also adopted this latter approach, preferring a more rigorous level of scrutiny of the evidence and justifications put before it by the state, to counter the effect of adopting

---

122 TAC (n 104 above) paras 46 & 80 - 81.
123 *Khosa v Minister of Social Development; Mahlaule v Minister of Social Development* 2004 6 SA 505 (CC) para 50.
a fairly weak interpretation of the right and a weak remedy. In *Grootboom*\(^{124}\) and *TAC*\(^{125}\), the weak remedy was partly attributable to the actual or perceived shift in state policy prior to the order being handed down, rendering a more forceful remedy unnecessary. It is only in *Khosa*\(^{126}\) that a direct, positive obligation is imposed upon the state.

More recently, the Constitutional Court has used the notion of meaningful engagement in its adjudication of evictions cases. In *Olivia Road*,\(^{127}\) the Court failed to engage with the ‘hard’ issues, preferring instead to refer the matter back to the parties to attempt to sort out the dispute between themselves. In doing so, the Court failed to engage with the substance of the attack on the constitutionality of the City’s housing policy and retreated to a consideration of procedural fairness. Similarly, in *Joe Slovo*,\(^{128}\) the Court ordered a massive eviction, even where the state had failed to engage meaningfully with those affected by its decision. In a context where the Constitution guarantees that no one will be evicted from a home without a court order made after considering all relevant circumstances, this failure is regrettable.

It is suggested that the concept of meaningful engagement, while rendering explicit an obligation to consult with those affected by administrative action, masks a further retreat by the Court in its differential treatment of socio-economic rights. The reasonableness test developed in *Grootboom*\(^{129}\) at its minimum, was an administrative law reasonableness review.\(^{130}\) *Joe Slovo* marks a preoccupation with procedural fairness, in an assessment of whether the eviction was just and equitable; at the same time the ‘laudable’ aims of a misconceived development were used to trump those very procedural fairness rights.

**4 Conclusion**

This article ends with some reflections regarding the use which this discussion of constitutional deference may have for an analysis of socio-economic rights adjudication. First, the principle of constitutional deference creates a framework within which to understand and critique the rhetoric of deference employed by the courts; second, it provides a perspective to understand what the courts are doing even when they do not use the language of deference; third, it allows for analysis of judicial

124 *Grootboom* (n 117 above) para 99.
125 *TAC* (n 104 above) paras 117 - 122.
126 *Khosa* (n 123 above).
127 *Occupiers of 51 Olivia Road, Berea Township, and 197 Main Street, Johannesburg v City of Johannesburg* 2008 3 SA 208 (CC) para 34.
128 *Residents of Joe Slovo Community, Western Cape v Thubelisha Homes* 2010 3 SA 454 (CC) para 7 order 16.
129 *Grootboom* (n 117 above).
reasoning across different types of cases and provides a framework for such an analysis; and fourth, and perhaps most importantly, it provides a framework within which judges themselves should understand their approach to a particular case. In this sense, a concept of constitutional deference is a call for greater transparency and self-reflection in judicial reasoning. There is no single or ‘correct’ approach to deference, or a correct deferential standard which the court should apply, just as there can never be a ‘correct’ understanding of democracy. Rather, what this approach to constitutional deference hopes to bring about is a greater consideration of the underpinning principles of constitutional deference and greater transparency and engagement on the part of the courts and commentators with these principles.
CHAPTER 14

RECLAIMING THE FRONTIER OF CONSTITUTIONAL DEFERENCE: MAZIBUKO V CITY OF JOHANNESBURG - A JURISPRUDENTIAL SETBACK*

Redson Edward Kapindu*

1 Introduction

The separation of powers of the various organs of state, and the concomitant concept of checks and balances as an integral part of its proper understanding, is a precondition for the observance of the rule of law and the effective guarantee of human rights.1 Seedorf and Sibanda state that the objective of the separation of powers is to curtail the exercise of political power and to prevent its abuse – meaning the violation of human rights – and that the underlying idea beneath this doctrine is the sceptical assessment that good governance is more likely when political power is distributed among different bodies or authorities.2 Thus, the various organs of the state must observe the scope of their province of authority and responsibility, and respect the province and authority of other state organs.

In Minister of Public Works v Kyalami Ridge Environmental Association (Kyalami case),3 the Constitutional Court stated that the Constitution makes provision for a separation of powers between the legislature, the executive and the judiciary, and that ‘this separation ordinarily implies that the legislature makes the laws, the executive implements them and the judiciary determines whether in the light of the Constitution and the law, conduct is lawful or unlawful’.4 The Court proceeded to hold that ‘though

---

* Lecturer, Department of Public Law, University of Johannesburg, and Deputy Director, South African Institute for Advanced Constitutional, Public, Human Rights and International Law (SAIFAC). An earlier version of this paper was presented at the SAIFAC ‘Is this seat taken?’ Colloquium held on 11 October 2009. I would like to thank the anonymous reviewer of this paper for his/her very valuable comments and encouragement.

3 2001 3 SA 1151 (CC).
4 Kyalami (n 3 above) para 37.
the separation prescribed by the Constitution is not absolute, and on occasions some overlapping of functions is permissible, action that is inconsistent with the separation demanded is invalid'.

In *Ferreira v Levin NO* (*Ferreira* case), Chaskalson P introduced the concept of pre-eminent domain, which means that while there will be instances where the functions of the principal organs of the state overlap, each organ must desist from intruding into a functional area that is pre-eminently within the domain of another. Commenting on the various functions of the principal organs of government, Chaskalson P observed that ‘at times these functions may overlap. But the terrains are in the main separate, and should be kept separate.’ In *Minister of Health v Treatment Action Campaign* (*TAC* case), the Constitutional Court affirmed this concept stating:

This Court has made it clear on more than one occasion that although there are no bright lines that separate the roles of the legislature, the executive and the courts from one another, there are certain matters that are pre-eminently within the domain of one or other of the arms of government and not the others. All arms of government should be sensitive to and respect this separation.

Within the framework of pre-eminent domain, Chaskalson P observed in *Ferreira* that ‘the protection of fundamental freedoms is pre-eminently a function of the court’.

This is an important proposition to which I shall return later in this discussion. At the same time, however, it is important for there to be a necessary and critical understanding among these organs that they are organs of one composite body, the state, and that they need to work in a spirit of unity and complement one another to achieve the common good of society as a whole. Though the idea of checks and balances seems on its face to carry connotations of inherent tension with each organ of the state checking on the excesses of the other, it is actually a useful and potent mechanism that gives effect to complementarity. This interplay of roles is essential in maintaining a constitutional democracy. At the core of constitutional democracy lies the notion of constitutional supremacy rather than parliamentary (or sometimes by the Constitution is not absolute, and on occasions some overlapping of functions is permissible, action that is inconsistent with the separation demanded is invalid'.

In *Ferreira v Levin NO* (*Ferreira* case), Chaskalson P introduced the concept of pre-eminent domain, which means that while there will be instances where the functions of the principal organs of the state overlap, each organ must desist from intruding into a functional area that is pre-eminently within the domain of another. Commenting on the various functions of the principal organs of government, Chaskalson P observed that ‘at times these functions may overlap. But the terrains are in the main separate, and should be kept separate.’ In *Minister of Health v Treatment Action Campaign* (*TAC* case), the Constitutional Court affirmed this concept stating:

This Court has made it clear on more than one occasion that although there are no bright lines that separate the roles of the legislature, the executive and the courts from one another, there are certain matters that are pre-eminently within the domain of one or other of the arms of government and not the others. All arms of government should be sensitive to and respect this separation.

Within the framework of pre-eminent domain, Chaskalson P observed in *Ferreira* that ‘the protection of fundamental freedoms is pre-eminently a function of the court’.

This is an important proposition to which I shall return later in this discussion. At the same time, however, it is important for there to be a necessary and critical understanding among these organs that they are organs of one composite body, the state, and that they need to work in a spirit of unity and complement one another to achieve the common good of society as a whole. Though the idea of checks and balances seems on its face to carry connotations of inherent tension with each organ of the state checking on the excesses of the other, it is actually a useful and potent mechanism that gives effect to complementarity. This interplay of roles is essential in maintaining a constitutional democracy. At the core of constitutional democracy lies the notion of constitutional supremacy rather than parliamentary (or sometimes by the Constitution is not absolute, and on occasions some overlapping of functions is permissible, action that is inconsistent with the separation demanded is invalid'.

In *Ferreira v Levin NO* (*Ferreira* case), Chaskalson P introduced the concept of pre-eminent domain, which means that while there will be instances where the functions of the principal organs of the state overlap, each organ must desist from intruding into a functional area that is pre-eminently within the domain of another. Commenting on the various functions of the principal organs of government, Chaskalson P observed that ‘at times these functions may overlap. But the terrains are in the main separate, and should be kept separate.’ In *Minister of Health v Treatment Action Campaign* (*TAC* case), the Constitutional Court affirmed this concept stating:

This Court has made it clear on more than one occasion that although there are no bright lines that separate the roles of the legislature, the executive and the courts from one another, there are certain matters that are pre-eminently within the domain of one or other of the arms of government and not the others. All arms of government should be sensitive to and respect this separation.

Within the framework of pre-eminent domain, Chaskalson P observed in *Ferreira* that ‘the protection of fundamental freedoms is pre-eminently a function of the court’.

This is an important proposition to which I shall return later in this discussion. At the same time, however, it is important for there to be a necessary and critical understanding among these organs that they are organs of one composite body, the state, and that they need to work in a spirit of unity and complement one another to achieve the common good of society as a whole. Though the idea of checks and balances seems on its face to carry connotations of inherent tension with each organ of the state checking on the excesses of the other, it is actually a useful and potent mechanism that gives effect to complementarity. This interplay of roles is essential in maintaining a constitutional democracy. At the core of constitutional democracy lies the notion of constitutional supremacy rather than parliamentary (or sometimes

5. *Kyalami* (n 3 above) para 37. Also on this point, see the cases of *Executive Council, of Western Cape Legislature v President of the Republic of South Africa* 1995 4 SA 877 (CC); 1995 10 BCLR 1289 (CC) para 55; *Ex-Parte Chairperson of the Assembly, In Re Certification of the Constitution of Republic of South Africa, 1996 1996 4 SA 744 (CC); 1996 10 BCLR 1253 (CC) para 111; *Ex-Parte Speaker of the Western Cape Provincial Legislature: In Re Certification of the Constitution of the Western Cape, 1997 1997 4 SA 795 (CC); 1997 9 BCLR 1167 (CC) para 63; *De Lange v Smuts NO* 1998 3 SA 785 (CC); 1998 7 BCLR 779 (CC) para 124; *South African Association of Personal Injury Lawyers v Heath* 2001 1 SA 883 (CC); 2001 1 BCLR 77 (CC) paras 23 - 26.


7. *Ferreira* (n 6 above) para 183.

8. 2002 5 SA 703 (CC).

9. *TAC* (n 8 above) para 98 (my emphasis).

10. *Ferreira* (n 6 above) para 183.
executive) supremacy. Whilst parliamentary supremacy is a direct offshoot of an understanding of democracy in the sense of ‘representative democracy’, constitutional democracy, as Roux states, is a purely descriptive term that is used to connote a system in which the people’s power to make collective decisions is constrained by a written constitution, or at least a received set of institutional practices that is regarded as being incapable of ordinary amendment. Roux argues that usually, but not necessarily, the power to decide whether the people, acting through the political branches, have deviated from the terms of the Constitution is vested in the judiciary, and also that usually but not necessarily, the judiciary exercises this power on the authority of the supreme law embodied in the bill of rights. Roux understands the term ‘constitutional democracy’ as the binary opposite of the term ‘parliamentary sovereignty’ which connotes a system in which the legislature has the final word in the event of inter-branch conflict over the constitutionality of a collective decision. It has been argued, quite correctly, that since the judiciary is the final arbiter on the constitutionality of laws or conduct in a system of constitutional supremacy and, indeed, since the courts make the final determination on the scope of their own powers, the courts must have mechanisms of self-restraint to prevent them from unnecessarily interfering with the functions of the other organs of government. They need to be fully cognisant of the need to allow an appropriate margin of discretion for the other state organs.

It is submitted that the notion of judicial deference towards other organs of government ought to be understood within the framework of constitutional democracy as eloquently expounded by Roux. Kirsty McLean’s paper reflects this understanding to a large extent. However, what seems to emerge from her paper is that she does not necessarily support or attack the nature of judicial deference that the Constitutional Court has adopted in socio-economic rights cases; and this seems to be deliberate as her project is only meant, it seems, to be a principled demonstration of the use of the post-apartheid understanding of constitutional deference in an analysis of human rights and particularly socio-economic rights adjudication.

This paper generally affirms McLean’s project in this regard, but wishes to go further and submit that the Constitutional Court has extended its notion of deference too far. The paper surveys a number of decisions on socio-economic rights by the Court, but it will focus in more depth on the Court’s recent decision in Mazibuko v City of Johannesburg. This is a landmark

---

11 Executive supremacy manifests itself in dictatorships where the head of state or government wields immense powers, sometimes issuing decrees or edicts that override laws passed by elected institutions. It is submitted that it would be a misnomer to refer to such a regime as reflecting parliamentary supremacy.
decision in the Constitutional Court’s socio-economic rights jurisprudence.

Among other things, the court not only pronounced on the obligations of the state with regard to the right of access to sufficient water in terms of section 27(1)(b) of the Constitution, but it also expressly pronounced on the ‘proper role of courts’ in determining the content of socio-economic rights in a constitutional democracy – an important exposition as it relates to the notion of judicial deference in these cases towards other organs of government. The court sought to settle the boundaries in view of the fact that, notwithstanding its modest stream of decisions on this cluster of rights thus far, the understanding of the nature of obligations imposed by the Constitution on the state, and the role of the courts in that regard, seems unstable.

The approach taken by the Constitutional Court in this case is highly deferential to the legislature and the executive. So deferential, this paper argues, as to signify a retreat from the socio-economic rights promised by the Constitution and from the role the court has played in promoting the much neglected socio-economic rights of the voiceless, powerless and marginalised members of South African society. The lines drawn by judicial deference in socio-economic rights cases, resting with the Mazibuko case, have relinquished the proper frontier of the judicial province and there is a need for the Court to reclaim its own territory.

This paper sees a useful analogy for this discourse in the approach taken in the jurisprudence of the International Court of Justice (ICJ) in relation to the exercise of its non-contentious (Advisory Opinion) jurisdiction. The paper suggests that the approach of the ICJ in this regard, as the principal judicial organ of the United Nations (UN), could provide a useful guide as to how courts should address the scope and meaning of the judicial function within the framework of separation of powers in socio-economic rights discourse.

2 Justifications for judicial deference in socio-economic rights cases: What are the grounds?

Although there is a growing trend to stress the universality, interdependence, interconnectedness and indivisibility of all human rights and that they should all be treated on the same footing and with the same emphasis, this apparent consensus masks a deep and enduring

---

15 Mazibuko (n 14 above) paras 46 & 57.
16 According to para 5 of the Vienna Declaration and Programme of Action (1993): ‘All human rights are universal, indivisible and interdependent and interrelated. The international community must treat human rights globally in a fair and equal manner, on the same footing, and with the same emphasis. While the significance of national and regional
disagreement over the proper status of socio-economic rights. Courts in many countries, including South Africa, as well as supra-national tribunals such as the African Commission on Human and Peoples’ Rights, have been quick to point out that there is a stark distinction between the nature and character of civil and political rights on the one hand and socio-economic rights on the other. Indeed, as the UN Committee on Economic, Social and Cultural Rights (CESCR) has stated, it is shocking:

that states and the international community as whole continue to tolerate all too often breaches of economic, social and cultural rights which, if they occurred in relation to civil and political rights, would provoke expressions of horror and outrage and would lead to concerted calls for immediate remedial action.

In the context of South Africa, as McLean suggests in her paper, the Constitutional Court has not dealt with the adjudication of socio-economic rights in the same way that it has treated civil and political rights. This is evidenced primarily through the adoption of a lower standard of review, and deferential remedies. The reasons for this deferential approach appear to be that the Court, despite acknowledging the equal status of socio-economic rights, retains a number of particularities and various historical, cultural and religious backgrounds must be borne in mind, it is the duty of states, regardless of their political, economic and cultural systems, to promote and protect all human rights and fundamental freedoms. In the Grootboom case, Yacoob J highlighted the interconnectedness of all human rights in the Constitution, holding that ‘the Constitution entrenches both civil and political rights and social and economic rights. All the rights in our Bill of Rights are inter-related and mutually supporting. There can be no doubt that human dignity, freedom and equality, the foundational values of our society, are denied those who have no food, clothing or shelter. Affording socio-economic rights to all people therefore enables them to the other rights enshrined in chapter 2’ (para 23). Mokgoro J similarly held in the Khosa case that ‘[t]he socio-economic rights in our Constitution are closely related to the founding values of human dignity, equality and freedom’ and that ‘the proposition that rights are inter-related and are all equally important, has immense human and practical significance in a society founded on these values’ (para 40).


In the Certification of the Constitution of the Republic of South Africa 1996 4 SA 744 (CC) (2nd Certification case), the Constitutional Court, using the International Covenant on Economic, Social and Cultural Rights as a framework for reference with regard to the guarantee of socio-economic rights under the Constitution, stated that these rights were ‘not fully enforceable immediately, each state party only binding itself “to the maximum of its available resources” to “achieving progressively the full realisation of the rights recognised in the present Covenant”’. In no way do we intend to denigrate the importance of advancing and securing such rights. We merely point out that their nature and enforceability differ materially from those of other rights (my emphasis).

In Purohit and Another v The Gambia (2003) AHRLR 96 (ACHPR 2003), the African Commission on Human and Peoples’ Rights read into art 16 of the African Charter on Human and Peoples’ Rights the obligation to take concrete and targeted steps (progressively) to ensure the full realisation of the right to health.

unarticulated and unexamined reservations to the adjudication of these rights.\textsuperscript{21}

Of course, to be fair to the Court, it has tried, to some extent, to articulate and examine some of its reasons for adopting a rather restrained approach in the adjudication of socio-economic rights cases. Two major arguments, one principled and the other pragmatic,\textsuperscript{22} and both well known in socio-economic rights discourse, have been advanced by the Court:

1. The principled argument is framed in terms of legitimacy, and claims that the judiciary lacks democratic legitimacy to deal with matters that have complex distributive and redistributive socio-economic effects, and that these have to be left to the decisions of those political organs of the state that are directly accountable to the electorate. The argument includes the understanding that making decisions on the distribution of resources in society primarily involves the exercise of a political rather than a judicial function. All in all, it is urged that deference in this regard mitigates what is perceived as the anti-democratic nature of judicial review;\textsuperscript{23}

2. The pragmatic argument is based on the premise of competence, and views the judiciary as institutionally incompetent to deal with questions that involve complex policy considerations that have budgetary or redistributive ramifications.

3 Clarifying the role of courts in adjudicating socio-economic rights

In the light of the foregoing discussion, a key question arises: What exactly is the judicial function and what is the role of the judiciary in adjudicating socio-economic rights cases supposed to be? Section 165(1) of the Constitution provides that the judicial authority of the Republic vests in the courts, but it does not specifically define what essentially constitutes the judicial function. The task of deciding this has been left to the courts themselves.

In \textit{South African Association of Personal Injury Lawyers v Heath},\textsuperscript{24} Chaskalson P described the role of the judiciary in the context of separation of powers and the enforcement of human rights. He held that the judiciary has a sensitive and crucial role to play in controlling the exercise of power and upholding the Bill of Rights.\textsuperscript{25} He stated that its function in this regard is to be ‘an independent arbiter of issues involving the division of powers

\textsuperscript{21} K McLean ‘Towards a framework for understanding constitutional deference’ in S Woolman & D Bilchitz (eds) \textit{Is this seat taken? Conversations at the Bar, the bench and the academy about the South African Constitution} (2012) 291.
\textsuperscript{23} As above.
\textsuperscript{24} 2001 1 SA 883 (CC).
\textsuperscript{25} Heath (n 24 above) para 46.
between the various spheres of government, and the legality of legislative and executive action measured against the Bill of Rights, and other provisions of the Constitution’. 26 This exercise, according to Chaskalson P, requires independence and involves the weighing up of information, the forming of an opinion based on information and the law, and the giving of a decision on the basis of a consideration of relevant information and the law. 27

As has been pointed out, the Ferreira case stressed that ‘the protection of fundamental freedoms is pre-eminently a function of the court’. 28 The Ferreira case concerned civil and political rights, but this statement is equally applicable to socio-economic rights. Thus, when a litigant approaches the Court because his or her constitutional right of access to sufficient water is being violated or is under threat of violation, it is pre-eminently a judicial task to determine whether the conduct or law in question constitutes such violation or threat of violation. Such a decision will involve the weighing up of information, the forming of an opinion based on the available information (or facts and the law, and the giving of a decision on the basis of a consideration of the relevant information (or facts) and the law. In the TAC case, the Constitutional Court provided a lucid statement of the role of courts with regard to socio-economic rights. It stated that:

The primary duty of courts is to the Constitution and the law, ‘which they must apply impartially and without fear, favour or prejudice’. The Constitution requires the state to ‘respect, protect, promote, and fulfil the rights in the Bill of Rights’. Where state policy is challenged as inconsistent with the Constitution, courts have to consider whether in formulating and implementing such policy the state has given effect to its constitutional obligations. If it should hold in any given case that the state has failed to do so, it is obliged by the Constitution to say so. In so far as that constitutes an intrusion into the domain of the executive, that is an intrusion mandated by the Constitution itself. 29

Thus, the Court seems to suggest that where a court delves into policy issues with a view to assessing the compatibility of such policies with the Constitution, whilst that might be viewed as an intrusion into an area that is not pre-eminently within its domain, it is an intrusion that is constitutionally mandated. One might call this a technical constitutional justification for the court to engage complex polycentric policy issues. The Court points out that the reason it would engage in such an exercise is to decide on ‘whether in formulating and implementing such policy the state has given effect to its constitutional obligations’. It is submitted that the Court cannot do this without defining the content of such obligations.

The question remains: Is it essentially inconsistent with the judicial function for the court to deal with policy issues? Some have argued that the most

26 Heath (n 24 above) para 26.
27 Heath (n 24 above) para 34.
28 n 7 above.
29 TAC (n 8 above) para 99.
effective realisation of socio-economic rights depends upon policy choices and
that it is precisely in choices of this nature where judges lack two essential
qualifications: expertise and political accountability.\textsuperscript{30} This argument
however does not seem unassailable.

Judges in common law jurisdictions have engaged in policy considerations
for a very long time. Richardson has argued, for instance, that judges make
law and are expected to make law and in doing so, they necessarily weigh
policy considerations.\textsuperscript{31}

Wade similarly contends that judges ‘are up to their necks in policy, as they
have been all through history’.\textsuperscript{32} Hunt argues that ‘judicial law making is not
incidental or peripheral to policy matters’. He argues that, on the contrary,
‘it has shaped concepts and principles with crucial policy content, such as the
law of negligence and rules of natural justice’.\textsuperscript{33} In \textit{Port Elizabeth Municipality
v Various Occupiers} (the \textit{Port Elizabeth Municipality} Case),\textsuperscript{34} Sachs J stated that:

\begin{quote}
The integrity of the rights-based vision of the Constitution is punctured when
governmental action augments rather than reduces denial of the claims of the
desperately poor to the basic elements of a decent existence. Hence the need for
special judicial control of a process that is both socially stressful and
potentially conflictual.\textsuperscript{35}
\end{quote}

Sachs J suggests that, firstly, the socio-economic rights guaranteed by the
Constitution provide the people with claims. This rights-based vision seeks to
reduce or avert the denial of these claims.

As I will later demonstrate, the \textit{Mazibuko} decision represents a
complete retreat from this position as the Court denies that the people have
such claims under the Constitution. This leaves very little of the rights-based
vision of the Constitution put forward by Sachs J. Secondly, the effective
guarantee of socio-economic rights frequently involves the distribution
and/or redistribution of resources, and in this case Sachs J requires ‘a
process that is both socially stressful and conflictual’ which, according to
her, is the very reason why the judiciary has to engage with these issues.
This addresses the competence concern in the sense that the courts are
seen as competent to deal with polycentric conflict-ridden issues that are
socially stressful. However, in the unanimous Court decision in \textit{Mazibuko},
which included Sachs J himself, retreated from this position and pushed
the frontiers of its deference towards other state organs much further.

\textsuperscript{30} E Mureinik, ‘Beyond a charter of luxuries: Economic rights in the Constitution’ (1992) 8
\textit{SAJHR} 464 467.
\textsuperscript{31} WS Richardson ‘Public interest litigation’ (1995) 3 \textit{Waikato LR} 1.
\textsuperscript{32} HWR Wade \textit{Constitutional fundamentals} (1989) 78.
\textsuperscript{34} 2005 1 SA 217 (CC).
\textsuperscript{35} \textit{Port Elizabeth Municipality} (n 35 above) para 18.
An innovative way in which courts, most notably in India, have dealt with the controversial question of their competence is to rely on expert bodies to study the situation and submit reports to the Court. Courts have also appointed commissions to propose remedial relief and monitor its implementation. Desai and Muralidhar submit that the use of commissions has enabled the court to check the facts as alleged by the petitioner and the state after a proper scrutiny without affecting its role as an adjudicator. Such a process is not unique to common law jurisdictions. It is not uncommon for courts to appoint assessors who are experts in relevant fields to assist the court to arrive at a proper, sound and informed decision. The Constitutional Court could adopt this measure to avoid excessive deference and the dilemma of having either to abdicate its judicial role or to usurp executive and legislative responsibilities.

To further appreciate the meaning and scope of the expression ‘judicial function’ lessons may be drawn from the approach taken by the International Court of Justice (ICJ) when it assumes jurisdiction in non-contentious cases. In Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, the ICJ stated that ‘the Court has recalled many times in the past that Article 65, paragraph 1, of its Statute, which provides that “The Court may give an advisory opinion …” should be interpreted to mean that the Court has a discretionary power to decline to give an advisory opinion even if the conditions of jurisdiction are met’.

This passage clearly shows that the ICJ has a lot of room for according deference to other organs of the UN such as the General Assembly and the Security Council. In requests made to the Court to provide Advisory Opinions, the issue as to whether the question posed is of a judicial or legal

---

37 As above.  
38 As above.  
39 Ferraz has argued that whilst ‘the South African reasonableness approach is seen as too deferential and abdicative of the judiciary’s role in protecting rights, the Brazilian individually enforceable rights approach is deemed too intrusive and usurpative of the prerogative of elected representatives to define how the limited resources of the state should be allocated among unlimited social needs’. He further argues that ‘the emerging co-operative constitutionalism theories, which try to apply the institutional dialogue theories of judicial review to social rights adjudication, do not solve the dilemma’ and that ‘they are currently largely procedural, and therefore liable to the same charges of abdication levelled at the reasonableness approach of the South African Constitutional Court’. But if they try to become more substantive, he urges, they will certainly attract the accusations of usurpation currently levelled at more assertive courts such as the Brazilian ones. He therefore concludes that the dilemma might be insoluble until either a more stable consensus emerges on what social and economic rights entail or the expectation that rights necessarily imply strong judicial remedies gradually wanes. See OLM Ferraz ‘Between usurpation and abdication? The right to health in the courts of Brazil and South Africa’ http://papers.ssrn.com (accessed 16 March 2010).  
41 Legality of the threat or use of nuclear weapons (n 40 above) para 14.
character, and not inconsistent with the Court’s judicial function, has consistently been raised by disputants. However, although article 65(1) of the ICJ Statute plainly allows plenty of room for deference towards other organs of the UN and other international actors and the discretion to refuse to assume jurisdiction, the ICJ has made it plain that although it has this discretion, the Court has never refused a request to provide an Advisory Opinion.42

The Court stresses that when a case includes political aspects as so often happens in the international arena, it does not serve to deprive the case of its character as raising a ‘legal question’ or to ‘deprive the Court of a competence expressly conferred on it by its Statute’.43 The ICJ has insisted that whatever its political aspects, the Court cannot refuse to admit the legal character of a question which invites it to discharge an essentially judicial task.44 In **Legality of the Threat or Use of Nuclear Weapons** case, the ICJ stated that its judicial function involves ascertaining the existence or otherwise of legal principles and applicable rules, and stating the law, even if, in stating and applying the law, the Court necessarily has to specify its scope and content and sometimes note its general trend.45 Critical to note is the point that in stating and applying the law, the Court necessarily has to specify its scope and content. It is thus submitted that if, for instance, an issue pertaining to the right of peoples to self-determination as provided for in article 55 of the UN Charter arose, within the exercise of its judicial function, the ICJ would have to define the content of that right in order to specify its scope.

In spite of being urged not to exercise its Advisory Opinion jurisdiction – based on such arguments as legitimacy, separation of powers among the organs of the UN, institutional competence, polycentric issues, the fact that the General Assembly or the Security Council have the requisite competence – the ICJ has declined to abdicate its judicial task of determining the legal questions brought before it, or to clarify the scope of the law, its general trend, or to spell out the legal consequences of the conduct of subjects of international law. Indeed, the ICJ has emphasised that as a principal organ of the UN, its decisions in such cases should not be seen as a usurpation of the roles of other principal organs of the UN, but rather as representing its participation in the activities of the UN. Hence, requests to render such decisions should not be refused.46 Thus, notwithstanding the very permissive words of article 65(1) of the ICJ statute for the Court

---

to defer in its processes to other actors, the ICJ has held very tenaciously to its judicial mandate, irrespective of the multi-polar socio-political character of the issues before it or the likely consequences of its decisions.

It is submitted that a useful analogy can be drawn between the role of the ICJ in the UN system and the role of the courts in the South African domestic sphere. First, the judiciary is a principal organ of government, just as the ICJ is the UN’s principal judicial organ, and when determining questions of constitutionality, including the enforcement of socio-economic rights, judicial decisions should be viewed as representing the judiciary’s participation in the activities of government just as the ICJ has stressed its judicial role as a form of participation in the work of the UN.

This therefore calls for active dialogue between the various branches of government over fundamental rights and freedoms. Gauging the propositions of both the Constitutional Court, as well as the ICJ, the essence of the judicial function is to be an independent arbiter, applying the Constitution and the law impartially and without fear, favour or prejudice. This involves ascertaining the existence or otherwise of legal principles and rules applicable to a given scenario, and specifying the scope and general trend of those principles and rules. In discharging this function, there may be an interplay of polycentric socio-economic factors, but as the ICJ noted in the Nuclear Weapons case, that does not per se detract from the judicial character of the issue before the Court as long as the Court is called upon to weigh issues of fact against the content of the law – in this case, constitutional guarantees of socio-economic rights. To this end it is submitted that the definition of the content of rights, whether they be civil and political rights or socio-economic rights, falls within the heartland of the judicial function.

As the Constitutional Court observed in the TAC case, where state policy is challenged as inconsistent with the Constitution, the court had to consider whether the state, in formulating and implementing such policy, had failed to give effect to its constitutional obligations and was obliged to point this out. Further, the Court emphasised that in so far as the Court’s order ‘constitutes an intrusion into the domain of the executive, that is an intrusion mandated by the Constitution itself and that, whilst such orders can affect the policies of government or other organs of state and have budgetary implications, ‘government is constitutionally bound to give effect to such orders whether or not they affect its policy and has to find the resources to do so’. Further, as the Court emphasised in Fose v Minister of Safety and Security, if it is necessary to do so, the courts may even have to

47 See Edwards (n 21 above) 866.
48 See sec 165(2) of the Constitution.
49 See Heath (n 22 above).
50 n 28 above.
51 As above.
52 As above.
fashion new remedies to secure the protection and enforcement of these all-
important rights.53

Courts therefore should not hasten to defer the definition of the content
and scope of socio-economic rights to other branches of the government
as this is an abdication of their own responsibility.

4 Exploring judicial deference in South African
socio-economic rights jurisprudence

As stated in the introduction, the Constitutional Court’s approach to the
enforcement of socio-economic rights has been unstable.

It first considered the subject in *Ex Parte Chairperson of the Constitutional
Assembly: In Re Certification of the Constitution of the Republic of South Africa (1st
Certification case)*,54 and responded lucidly to the objection that the
inclusion of socio-economic rights in the Constitution would breach the
principle of separation of powers as follows:

It is true that the inclusion of socio-economic rights may result in courts
making orders which have direct implications for budgetary matters. However,
even when a court enforces civil and political rights such as equality, freedom
of speech and the right to a fair trial, the order it makes will often have such
implications. A court may require the provision of legal aid, or the extension of
state benefits to a class of people who formerly were not beneficiaries of such
benefits. In our view it cannot be said that by including socio-economic rights
within a bill of rights, a task is conferred upon the courts so different from that
ordinarily conferred upon them by a bill of rights that it results in a breach of
the separation of powers.55

This early case injected hope that socio-economic rights would soon be placed
on an equal footing with civil and political rights in the context of
justiciability, but hope began to wane after the remarks of Chaskalson P in
*Ferreira*:56

Whether or not there should be regulation and redistribution is essentially a
political question which falls within the domain of the legislature and not the
court. It is not for the courts to approve or disapprove of such policies. What
the courts must ensure is that the implementation of any political decision to
undertake such policies conforms with the Constitution.57

Chaskalson’s position is that the Court may not review the policy adopted
by the legislature (or executive) or the political decision to undertake a policy,

53 1997 3 SA 786 (CC) para 19.
54 1996 4 SA 744; 1996 10 BCLR 1253 (CC).
55 *First Certification judgment* (n 54 above) 78.
56 *Ferreira* (n 6 above) para 180.
but it may review the implementation of a political decision to undertake such policy. This shows a great measure of deference.

Soobramoney v Minister of Health, KwaZulu-Natal (Soobramoney case)\(^{58}\) was the first decision in which the principal issue before the Constitutional Court rested on the interpretation and enforcement of socio-economic rights, and the Court observed that it was primarily the duty of the provincial administration to make decisions about the funding that should be made available for health care and how such funds should be spent. These choices, the Court observed, involved difficult decisions to be taken at the political level in fixing the health budget, and at the functional level in deciding upon the priorities to be met. It then held that a court ought to be slow to interfere with rational decisions taken in good faith by the political organs and authorities whose responsibility it is to deal with such matters.\(^{59}\)

Thus, the approach of the Court in Soobramoney seems consistent with the remarks, albeit obiter, by Chaskalson P in the Ferreira case. The Court in the Soobramoney case adopted rationality as the standard of review in determining the margin of discretion for political authorities that is beyond judicial scrutiny. As long as the decision taken by the political authorities with regard to the allocation or distribution of resources was rational, the courts could not intervene.

The rationality standard adopted in Soobramoney was later refined into the reasonableness review in Government of South Africa v Grootboom (Grootboom case),\(^{60}\) which is a more substantive and less deferential standard.\(^{61}\) Mbazira rightly suggests however, that while the reasonableness approach is more rigorous than the rationality test adopted in Soobramoney, it still has a significant element of judicial deference, and remains respectful of the [representative] democratic decision-making process and the limited nature of public resources, while simultaneously bestowing special deliberative attention upon those whose minimal needs are not being met.\(^{62}\)

As it has evolved thus far, the reasonableness test is rather a shy approach. It seeks to ensure that those in desperate need are included in policy making, especially those that further socio-economic rights, and to ensure that those policies do not result only in statistical progress but treat each human being as a person deserving of care and concern,\(^{63}\) but the reasonableness test has failed to define the content of particular rights. In adopting this test, the

\(^{58}\) 1998 1 SA 765 (CC).

\(^{59}\) Soobramoney (n 58 above) para 29.

\(^{60}\) 2000 11 BCLR 1169 (CC).


\(^{62}\) Mbazira (n 61 above) 11.

\(^{63}\) Grootboom (n 60 above) para 44.
Court avoids adopting the minimum core content approach, whilst in essence embracing the basic philosophy behind the same.64

Following on the Grootboom case was the TAC case where it was held that the courts are not institutionally equipped to make the wide-ranging factual and political enquiries necessary to determine what the minimum-core standards should be, and further, that they are ill-suited to adjudicate upon issues where court orders could have multiple social and economic consequences for the community. The Court thus held that the ‘Constitution contemplates a rather restrained and focused role for the courts’65 and specifically rejected the applicability of the minimum core concept in South Africa. However, the TAC case is quite ironic in some respects as it proceeded, firstly, to arrive at a decision that had to do with the distribution of resources. A specific mandatory order was granted obliging government to distribute resources – in this case the essential drug nevirapine that helps to prevent mother to child transmission of HIV. Secondly, the Court made very interesting pronouncements that seem at odds with the ultimate position it took. As stated earlier, the Court held that it was proper for the Court, in appropriate socio-economic rights cases, to make orders that can affect the policies of government or other organs of state and which may have budgetary implications. It stated:

‘Government is constitutionally bound to give effect to such orders whether or not they affect its policy and has to find the resources to do so.’66 Thus, the Court here suggests that, as long as it is constitutionally mandated to do so, in this cluster of rights a Court may make an order that obliges the other branches of Government to identify and encumber the necessary resources. This is a telling illustration of the unsettled position the Constitutional Court has taken with regard to these cases.

In Khosa v Minister of Social Development (Khosa case),67 the Court seemed to depart from the deferential approach of refraining from expressing views on the budgets or implications of decisions taken, or policies adopted, by the political organs of the state on the premise that they are better institutionally equipped to do so. The question before the Court was whether the exclusion of permanent residents from accessing social grants available to citizens was unconstitutional in terms of section 27(1)(c) of the Constitution. One of the arguments advanced by the state was that including permanent residents in the social security scheme would impose an impermissibly high financial burden on the state.68 The Court confronted the budgetary figures placed before it directly. The Court carefully looked at, analysed and interpreted the figures and concluded that they did not ‘support the

64 R Kapindu From the global to the local: The role of international law in the enforcement of socio-economic rights in South Africa (2009) 46.
65 As above.
66 As above.
67 2004 6 SA 505 (CC).
68 Khosa (n 67 above) para 60.
contention that there [would] be a huge cost in making provision for permanent residents’, and concluded that ‘the cost of including permanent residents in the system will be only a small proportion of the total cost’.

The Court thus held that the exclusion of permanent residents from social security benefits was unconstitutional and Government was required to include them in the scheme.

It is submitted that by examining actual figures, interpreting them and in significant measure basing its decision on such interpretation, the court was implicitly, if unintentionally, refuting its earlier argument in the TAC case that it was not institutionally equipped to make such analyses. This further bolsters the argument that jurisprudence in this area has been in a state of flux.

In Residents of Joe Slovo Community, Western Cape v Thubelisha Homes, the Court also raised the difficulty of balancing competing interests as a basis for adopting a deferential approach towards the executive authorities. Chenwi and Tissington urge that the Court failed to properly assess the reasonableness of the government’s policy choices, and that the Court displayed a particularly deferential attitude to the government. They rue the fact that the Court allowed the eviction of a relatively large community based on a government project that had been implemented without proper consultation and did not make provision for affordable housing for the intended beneficiaries. Moseneke DCJ stated in his judgment that if the applicants were not to benefit from the Joe Slovo housing development scheme, then their eviction would result in them being ‘sacrificial lambs to the grandiose national scheme to end informal settlements’. Upon a closer examination of the case, Chenwi and Tissington take the view that the Joe Slovo residents are indeed ‘sacrificial lambs’ in the state’s quest to ‘eradicate’ informal settlements.

Liebenberg’s critique of the judgment highlighted that the decision in Joe Slovo affected about 20 000 residents and represented the first large-scale judicially-sanctioned housing eviction since the end of apartheid. This is not meant to suggest that the Constitutional Court is acquiring an insensitivity reminiscent of apartheid era courts, but rather that this criticism should sound as a ‘wake-up call’ to the Court. It is in danger of relinquishing its position as the ultimate guardian and protector of the socio-economic guarantees of the Constitution and of showing too much

---

69 Khosa (n 67 above) para 62.
70 As above.
72 L Chenwi & K Tissington ‘“Sacrificial lambs” in the quest to eradicate informal settlements: The plight of Joe Slovo residents’ (2009) 10/3 ESR Review.
73 Joe Slovo (n 71 above) 138.
74 Chenwi & Tissington (n 72 above) 23.
deference towards other branches of government. Chenwi and Tissington conclude their paper with the following postscript:

On 24 August 2009, the Constitutional Court reportedly quietly issued an order suspending the evictions ‘until further notice’. The eviction order was suspended after the Western Cape provincial Minister of Housing, Bonginkosi Madikizela, submitted a report to the Court stating that he had “grave concerns” that the “massive relocation” might end up costing more than it would to upgrade Joe Slovo’. He also raised concerns about the absence of a plan regarding those who would not be accommodated in the new housing in Joe Slovo, since the houses would not be enough, and they would therefore be left behind in TRAs. The suspension of the eviction order has been welcomed by the Joe Slovo residents.76

It is interesting that the concerns about the lack of planning for those who would not be accommodated in the new housing scheme, and about the history of broken promises, were raised in argument before the Court by the applicants and amici curiae. What is deeply worrying is that the Court was unwilling to make the necessary order on the basis of these submissions, and would address these concerns only when it received the confirmatory provincial ministerial report on those very same concerns. This shows a deep level of deference towards other organs, particularly the executive.

In the Mazibuko case, the Court, in a unanimous decision, came out even more clearly on what it perceives as the justification for its deferential approach. It is clear from the text of the decision that the Court purposely intended the decision to provide a general guide on how socio-economic rights cases are to be prosecuted and adjudicated generally and this underscores the importance of the decision. The Court has said that the ‘reasons are essentially twofold. The first reason arises from the text of the Constitution and the second from an understanding of the proper role of courts in our constitutional democracy’.77 This role of the Court, as McLean has ably articulated in her paper, and as explained above, is understood from two angles: legitimacy and institutional competence.

With respect to the first (textual) reason, the Court in the Mazibuko case argues that section 27(1) and (2) must be read together to delineate the scope of the positive obligation imposed upon the state to provide access to sufficient water, and that this obligation requires the state to take reasonable legislative and other measures to progressively achieve the right of access to sufficient water within available resources. The Court stresses that this obligation ‘does not confer a right to claim sufficient water from the state immediately’.78 The absence of such a right then, it appears, is used by the Court to justify leaving a large deferential space for the
legislature and the executive (in particular) to determine the precise contours of the enjoyment of the rights guaranteed under section 27(1) at any given time.

An analysis of the Court’s previous jurisprudence however seems to suggest that when the right of access to sufficient water under section 27(1)(b) is deconstructed (just as with the rest of the other rights provided for under section 27(1)), there are some aspects of those guarantees that guarantee immediate claims and hence impose immediate obligations. In the *Grootboom* case, the Court held that the state is under an obligation to provide immediate relief that would fulfil certain minimum standards for those in desperate need. In the context of housing, the Court held that the requisite standards, short of adequate housing, that would meet this test were durability, habitability and stability. Although it is arguable that the situation in the *Mazibuko* case, on its specific facts, was different in that there was nothing on record to show that there were people who, by reason of the policies adopted by the City of Johannesburg, were left in a state of desperation without any potable water, the problem is that the decision, as has been the trend in the socio-economic rights jurisprudence of the Court, does not seek to deconstruct the content of the right of access to sufficient water. The jurisprudence focuses on determining the obligations of the state. As Bilchitz has argued, the approach of the Court suggests that the content of the right itself is to be determined by taking account of available resources and comes close to viewing socio-economic rights under the Constitution as providing a right to reasonable government action. However, if the state is under an obligation to make immediate provision for those in desperate need, as the Court acknowledged in *Grootboom*, then correlative, this implies that under sections 26 and 27 of the Constitution, everyone has a right to the immediate provision of the basic necessities of life, including water, when in desperate need. After all, the Court in *Mazibuko* begins with an acknowledgment that ‘water is life’ and that ‘without it, we will die’.

An objection can indeed be raised that the approach of the Court thus far, most prominently exemplified by the *Mazibuko* decision, of first identifying the obligation of the state and then determining whether the obligation gives rise to an immediate right to claim a right, access to sufficient water in this case, is rather lopsided. The preferred approach should be to look at the right in section 27(1) and then ask whether the right gives rise to such an immediate obligation, or an immediate obligation of any other form, regard being had to section 27(2). The integrity of the rights-based vision of the Constitution that Sachs J talked about in the *Port*

---

79 See *Grootboom* (n 60 above) para 52.  
81 See *Grootboom* (n 60 above) para 44.  
82 *Mazibuko* (n 14 above) para 1.
Elizabeth Municipality Case, it is submitted, would only be preserved by adopting an approach that proceeds from an examination of the right and then to a determination of the concomitant obligations, and not vice versa. If the primary focus is on the nature of the obligation so as to determine the contours of the right, then that also suggests that the court’s primary focus is on the duty-bearer rather than the claim holder, and that is problematic because it is scarcely a rights-based approach. A rights-based approach ‘identifies right(s) holders and their entitlements and corresponding duty bearers and their obligations, and works towards strengthening the capacities of rights-holders to make their claims and of duty bearers to meet their obligations’. In the succinct words of Sachs J, ‘the integrity of the rights-based vision of the Constitution is punctured when governmental action augments rather than reduces denial of the claims of the desperately poor to the basic elements of a decent existence’. The Court must begin to engage and develop the content of the socio-economic rights under the Constitution, instead of focusing only on the development of the content of the socio-economic obligations of the state.

The second reason for adopting a deferential approach in Mazibuko was premised on what the Court termed the proper role of courts in a constitutional democracy. The Court stated that ‘ordinarily it is institutionally inappropriate for a court to determine precisely what the achievement of any social and economic right entails and what steps government should take to ensure the progressive realisation of the right’. The Court argues that this is a matter for what it calls the ‘democratic’ institutions of government, the legislature and the executive, that are best placed to investigate social conditions in the light of available budgets and to determine what targets are achievable in relation to these rights. The Court urges that it is indeed ‘desirable’ that this be so for the sake of democratic accountability since it is the programmes and promises of these organs of the state ‘that are subjected to democratic popular choice’. The Court held that ‘[n]ational government should set the targets it wishes to achieve in respect of social and economic rights clearly’, in light of the constitutional founding values of accountability, responsiveness and openness. In so doing, the Court argues, citizens are enabled to hold government politically accountable if the standard is not achieved and that it also empowers citizens to hold government accountable through legal challenge if the standard is unreasonable. In this way, the Court urges, socio-


84 Proceeding from the understanding of constitutional democracy as stated by Roux above, it is submitted that this terminology is incorrect. The Court is also a democratic institution of the state. Any other sense would represent a retreat from the idea of constitutional supremacy and pave the way for a gradual return towards parliamentary supremacy.

85 Mazibuko (n 14 above) para 61.
economic rights entrenched in the Constitution may contribute to deepening democracy.

Thus what emerges is that the Court holds that the determination of the content of socio-economic rights is within the province of the legislature and the executive, not the judiciary. The idea of the indivisibility and interconnectedness of all human rights comes under severe strain here. For instance, the Court suggests that government will be held to account through the ballot box in respect of the realisation of socio-economic rights. This seems to be at odds with the approach the Court adopted in *S v Makwanyane* where, in an oft-cited passage, Chaskalson P stated:

> Public opinion may have some relevance to the enquiry, but in itself, it is no substitute for the duty vested in the Courts to interpret the Constitution and to uphold its provisions without fear or favour. If public opinion were to be decisive there would be no need for constitutional adjudication. The protection of rights could then be left to Parliament, which has a mandate from the public, and is answerable to the public for the way its mandate is exercised, but this would be a return to parliamentary sovereignty, and a retreat from the new legal order established by the 1993 Constitution … The very reason for establishing the new legal order, and for vesting the power of judicial review of all legislation in the courts, was to protect the rights of minorities and others who cannot protect their rights adequately through the democratic process.86

In the *Mazibuko* case, the Court clearly states the general principle that the extent to which socio-economic rights are to be enjoyed should be left to the legislature and the executive as these bodies have the electorate’s mandate and are answerable to the electorate for the way they exercise that mandate. This conclusion is a departure from Chaskalson’s proposition in *Makwanyane* above and it is unclear why the Court did not consider the *Makwanyane* decision. Perhaps it was considered to be a decision on civil rights, but when viewed in context, the application of Chaskalson’s dictum above is not confined to civil and political rights; it is of general application in respect of all rights. The *dictum* is also relevant to the additional protection of the rights of minorities and others who cannot protect their rights through the democratic process. It is noteworthy here that in the *Mazibuko* decision, the Court focuses on the claims of ‘citizens’ and the empowerment of ‘citizens to hold government accountable’. Previously, in *Khosa v Minister of Social Development (Khosa case)*, the Court referred to the significance of the fact that the rights under section 27(1) of the Constitution are guaranteed for ‘everyone’. Mokgoro J stated:

> The socio-economic rights in sections 26 and 27 of the Constitution are conferred on ‘everyone’ by subsection (1) in each of those sections. In contrast, the state’s obligations in respect of access to land apply only to citizens. Whether the right in section 27 is confined to citizens only or extends to a broader class of persons therefore depends on the interpretation of the

86 *S v Makwanyane* 1995 4 SA 391 para 88 (my emphasis).
word ‘everyone’ in that section … This Court has adopted a purposive approach to the interpretation of rights. Given that the Constitution expressly provides that the Bill of Rights enshrines the rights of ‘all people in our country’, and in the absence of any indication that the section 27(1) right is to be restricted to citizens as in other provisions in the Bill of Rights, the word ‘everyone’ in this section cannot be construed as referring only to ‘citizens’.87

Again, the Mazibuko decision does not attempt to refer to its previous decision in Khosa, and how its analysis on the nature of obligations imposed by section 27(1) of the Constitution and the role courts have to play vis-à-vis the political process, relates to the Makwanyane and Khosa decisions. It is submitted that in view of the positions adopted in these two previous decisions, it would be very difficult to justify the overly deferential approach to the role of political institutions in determining the content of the rights under section 27(1) in view of the fact that the section applies to everyone who is subject to the jurisdiction of South Africa, including those who cannot protect their rights adequately through the (popular) democratic process.

Finally, it is submitted that this approach is at odds with the judicial function of the Court vis-à-vis the Constitution. Section 167(3)(a) of the Constitution states that the Constitutional Court is the highest court in all constitutional matters, and section 167(7) provides that a constitutional matter includes any issue involving the interpretation, protection or enforcement of the Constitution. Put differently, any matter that involves the interpretation, protection and enforcement of the Constitution falls within the jurisdictional province of the Constitutional Court. Thus the court can only justify its position that the determination of ‘precisely what the achievement of any social and economic right entails’ is a matter for the legislature and the executive and not the courts if it can show that this issue does not involve the interpretation, protection and enforcement of the Constitution. If it does involve these matters, and there is a dispute, then the Court must interpret the right and this would include defining the scope of the right.

On the facts of the Mazibuko decision, it is submitted that the issues before the Court did indeed trigger all three requisite aspects: the interpretation, the protection and the enforcement of constitutional rights. The Court should not have said that the task of determining what a right entails, which is an interpretative task, lies anywhere other than with the Constitutional Court.

Thus, the implications of the Mazibuko decision, in so far as the case is held up as a guide for the approach the courts should take in socio-economic rights cases, are significant and profound. Leaving the question of the interpretation of the scope of rights to non-judicial branches of the

87 Khosa (n 60 above) paras 46 - 47.
government will, among other things, put those people that do not have a ‘significant’ enough voice to ensure their concerns will be heard in a particularly vulnerable position in so far as the protection and enforcement of these rights are concerned.

5 Conclusion

Under the Constitution, it is the responsibility of the judiciary to interpret and apply the law so as to advance socio-economic rights. Consequently, their decisions will have profound socio-economic consequences, but this should not cause the court to fear the vigilant enforcement of these rights. It has a duty to impartially and fearlessly interpret and apply the law with due respect and appreciation for the separation of powers. However, the Court should not be too amenable and defer decisions that legitimately belong within the boundaries of the judicial function to other organs, on the basis of orthodox arguments that were used previously to deny altogether the justiciability of socio-economic rights. It will prove unsustainable for the Constitutional Court to counsel, as it did in Mazibuko, that as the judiciary is not a democratic institution of the state, it falls to parliament and the executive to determine the content of the constitutional right at issue.

This is a minimalist conception of representative democracy. Within the context of constitutional democracy, it is submitted that the judiciary is very much a democratic institution of the state. South Africa’s democracy would be meaningless in the absence of a strong, independent and effective judiciary. As the Court itself observed in the TAC case, if some of its decisions are deemed to be an intrusion into the provinces of other organs, such intrusion is legitimate as long as it is sanctioned by a democratic and legitimate constitution. It is therefore submitted that the decision of the Constitutional Court in the Mazibuko case, given that it sets out to guide the adjudication of future decisions involving socio-economic rights, represents a significant setback in the pursuit of a more progressive regime and jurisprudence for the protection of socio-economic rights in South Africa. This paper contends that the Court’s active involvement in socio-economic matters represents its participation in the activities of government as one of its principal organs. As such, the Constitutional Court is urged not accede too readily to invitations for constitutional deference and yet not to usurp the functions of other state organs. The fundamental point is that the responsibility to define the content and scope of constitutional rights, whether they are civil and political rights or socio-economic rights, remains primarily that of the judiciary.
ON THE COMMON SAYING ‘WHAT’S TRUE IN GOLF IS TRUE IN LAW’: THE RELATIONSHIP BETWEEN THEORY AND PRACTICE ACROSS FORMS OF LIFE*

Stu Woolman**

Ya know, it’s a work in progress. Sean has a different theory about the mechanics of the swing. It takes some time – lots, and lots of practice – until the theory becomes second nature. It’s working, and every day I understand what Sean is saying, the theory, the mechanics, better and better. It’s slow, and I’m grinding, but I’m starting to see the results.

Tiger Woods (September 2010)

A picture [has] held us captive. And we could not get outside it, for it lay in our language and language seemed to repeat it to us inexorably (para 115) ... What we do is to bring words back from their metaphysical to their everyday use. (para 116) ... Now, however, let us suppose that after some efforts on the teacher’s part [the student] continues the series correctly, that is, as we do it. So now we can say he has mastered the system – But how far need he continue the series for us to have the right to say that? Clearly you cannot state a limit here. (para 145) ... The grammar of the word ‘knows’ is evidently closely related to that of ‘can’, ‘is able to’. But also closely related to that of ‘understands’. (‘Mastery’ of a technique) (para 150).

Ludwig Wittgenstein Philosophical investigations (1945)

[That shared practices are constitutive of our being] implies that the world is already given as the common world. It is not the case that there are first individual subjects which are at any given time have their own world; and that [through] [we then] the task of putting them together, by virtue of some sort of arrangement, ... [we then] would have a common world. We take pleasure ... as one takes pleasure, we speak ... about something as one speaks.

Martin Heidegger Lectures on the history of the concept of time (1926)

* I want to thank Deputy Chief Justice Dikgang Moseneke for initiating this conversation. See, eg, Moseneke ‘The relationship between legal theory and legal practice’ Paper delivered at the SAIFAC Is this seat taken? Conference (11 October 2009) (notes on file with author). I have, it goes without saying, benefitted immensely from the spirited reply of my colleagues Associate Professor and Director of SAIFAC, David Bilchitz and SAIFAC Researcher Juha Tuovinen.

** Professor and Elizabeth Bradley Chair of Ethics, Governance and Sustainable Development, University of the Witwatersrand; Academic Director, South African Institute for Advanced Constitutional, Public Human Rights and International Law (SAIFAC).
Chapter 15

Introduction

The way the vast majority of us (legal academics in South Africa) think about the relationships between legal theory and legal practice is dramatically out of step with what the majority of (non-legal) philosophers, neuroscientists, cognitive psychologists, economists, and other behavioural researchers have to say about the relationship between theory and practice. One consequence of these erroneous views is that the manner in which the majority of us understand ‘theory’ – as a metaphysical term – is sharply at odds with how things actually are. We replicate similar kinds of errors when we think about how various forms of life, ways of being in the world or social practices are constructed and how change actually occurs within such practices. Once again, epistemological fallacies with regard the role of legal theory have the consequence of leading us to reify legal theory in a manner that closer inspection shows to be unwarranted. This second error (‘the reification of theory’) leads many a legal academic to offer theoretical critiques at odds with what we know about the human condition and of limited value with regards to the reconstruction of the form of life we call ‘law’.

Advocates and attorneys in South Africa know something legal academics often forget: Practice and theory work hand in glove. For the most part, theory is understood as therapy for practice and they endorse the Wittgensteinian proposition that [legal] ‘philosophy is a battle against the bewitchment of our intelligence by means of language’.\(^1\) What they are aware of – consciously or not – is that we often mistakenly speak of the social practices, endowments and associations that make up our lives as if we were largely free to choose them or make them up as we go along. We are not. As Walzer argues, there is a ‘radical givenness’ to our social world and the practices that make it up.\(^2\) What he means, in short, is that most of the practices that make up our social life are involuntary. We do not choose our family. We generally do not choose our race or religion or ethnicity or nationality or class or citizenship. More to the point – we rarely choose or even influence the legal system that controls part of our social life. Moreover, even when we appear to have the space to exercise choice, we rarely create the practices within which we participate. The vast majority of our practices and forms of life are already there, culturally determined entities that pre-date our existence or, at the very least, our recognition of the need for them. Finally, even when we overcome inertia and do create some new practice (and let me not be understood to underestimate the value of such overcoming and creativity), the very structure and style of the practice is almost invariably based upon an existing rubric. So gay marriages may be truly new – but marriage itself is a

---

long-standing publicly-recognised institution for carrying on intimate or familial relationships. Even in times of radical transformation, reiteration and mimicry of existing social practices is the norm.

A reader might want to stop and say, ‘But wait, isn’t Tiger Woods talking about how theory is altering his practice in the quote above?’ Yes. That’s true. But to a point. And a very limited one. When the greatest golfer ever to play the game talks about theory and change, he is talking about what it means for someone who has already mastered the finest points of the game (after over 35 years of practice) to make incremental modifications designed to improve his ball striking. Everyone knows that you can read the entire oeuvre of golf literature, and never get the ball in the air. Coaches are for Woods what Wittgenstein’s philosophers are to other discursive practices – aids against bewitchment. It might feel to Woods, at any given moment, that he is doing things correctly. But it takes the trained and patient eye of another (his coach) to spot the tiny variations that make all the difference between a good swing and a bad one.

While I do not wish to diminish the contributions of fellow members of South Africa’s legal academy, it seems to me that we – and I include myself – often get the relationship between theory and practice back to front. To some extent, our betters and predecessors, such as Ronald Dworkin, are responsible for this mistake. From Taking rights seriously, to Law’s empire to Law’s domain, Dworkin has laid out modernity’s deepest and broadest theory of law, generally, and of constitutional law, in particular. General bewitchment by Dworkin’s project hardly comes as a surprise.

The first problem, as Dewey pointed out, is that Dworkin (and those legal philosophers who emulate him) have ‘arrogated to [themselves] the office of demonstrating the existence of a transcendent, absolute or inner reality and revealing to man the nature and features of this ultimate and higher reality’. As Wittgenstein and Heidegger note (as per the quotes above), no such transcendent, higher reality exists. The practice, and our ability to master it, comes first.

Only subsequent to mastery may we indulge in critique.

---

3 That a constitutional theory must be a theory of ‘everything’ may still attract a certain following, but my sense is that most serious contemporary constitutional law scholars would now reject that proposition. See LH Tribe Constitutional choices (1988) (‘My reply to the grim metamorphosis of constitutional argument ... into instrumental calculations of utility or, as pseudo-scientific calibrations ... in which the “costs” ... are supposedly “balanced” against the “benefits” ... is not to propose an alternative method ... My reply is to question all formulas as concealing the constitutional choices that we must make – and that we cannot responsibly pretend to derive from any neutral method.’)

4 J Dewey Reconstruction in philosophy (1920) 23.

5 Wittgenstein (n 2 above) (para 150) (‘mastery’ of a technique.)
The second problem is the response of the legal philosopher to the recognition that the inheritance of a common world does not forestall error. (Nor need it!) For while we are indeed endowed with a broad array of mutually-supporting beliefs, theories, conditions and standards, we cannot claim that this inheritance is a seamless whole, that all our beliefs are true. Only a hopelessly naïve epistemologist would entertain such a proposition. As Donald Davidson notes: ‘Error is what gives belief its point.’

Across the broad domain of social practices, however, we can claim to possess ‘endless true beliefs’. Of course, it does not follow that the practices themselves are entirely coherent. Our practices, traditions, institutions and truth propositions are, as Cohen writes:

... the result, not of legislative design by a single acting on behalf of a coherent system of values, but of conflicts among individuals acting on behalf of diverse values and ambitions. And unlike the produce of a supreme legislative design, the outcomes of such a history are not likely to be a set of coherent social practices that completely conform to a single scheme of values.

The problem with the theorist who reifies theory is that she mistakes these ‘faults, fizzes and heterogeneous layers’ within a practice as a problem with the practice – and its usefulness – as a whole. Take for example, Einstein’s contributions to the birth of both relativity theory and quantum mechanics. No working physicist argues that Einstein’s views were wrong – in the main – about both bases for modern physics. And yet, on particular points, Einstein’s views about both have been proven incorrect.

The third form of bewitchment, as Wittgenstein would have it in many of his ‘just so’ stories, is that we first form theories and then test these theories against experience. (Not so says Wittgenstein.) It is essential that we get our order of priority straight. Once a practice is established (through trial and error), we might wish, upon reflection, to test its assumptions through experiments that do or do not confirm aspects of the practice’s usefulness. But that is not how those who reify theory operate. Think again about Wittgenstein’s various *aperçu*:

Whence comes the idea that the beginning of a series is a visible section of rails invisibly laid to infinity? And infinitely long rails correspond to the unlimited application of a rule (para 218) ... But if something of this sort were the case, how would it help? No; my description only made sense if it was to be understood symbolically – I should have said: This is how it strikes me ... The word ‘agreement’ and the word ‘rule’ are closely related to one another, they are cousins. If I teach anyone the use of the one word, he learns the use of the other with it (para 224) ... and they agree in the language they use. That is not agreement in opinions but in a form of life (para 241) ... If language is to be a

---

7 As above.
means of communication there must be agreement not only in definitions but also (queer as this may sound) in judgments (para 250).

What Wittgenstein is saying is (a) that we already have the material as hand to arrive at verifiable truth propositions about the world; (b) that most of these propositions are shared; and (c) that most of them are true. Only at the margins, once we have aggressively learned all there is to be learned from one another, do our differences have any meaningful bite.9

The point of this initial exercise is to begin to establish the truth of these last few propositions across various forms of life, ways of being in the world, or other domains of human endeavour. But of course what brings us here is our mutual interest in the ontological status of ‘the law’. The rest of this article spins these initial observations into a concrete set of conclusions by making three relatively discrete moves. In part 2, I contend that what many readers would contend is a non-cognitive, non-normative practice – golf – is as cognitive and normative a practice as law (or, if you prefer, ‘law’ is as non-cognitive and non-normative, in the main, as golf.) The purpose here is to demonstrate the hand-in-glove fit between a practice and any theory about that practice in most forms of life. In part 3, I comment on a classic, well-known clash between two legal philosophers, Ronald Dworkin and Stanley Fish. In their game of GOTCHA!, the former grants theory primacy across the broad domain of human practices (including baseball), while the latter decries theory as epiphenomenal in virtually all such practices. Both are wrong – and wrong in a manner that has clear consequences for the practice of and theories about law. In part 4, I offer an account of the relationship between legal theory and legal practice that better fits the model of theory and practice adumbrated above. In this chastened account of legal theory, a modest account grounded in our recognition of the infinite amount about the world we do not know, practice (and direct engagement with the world) is accorded its proper place. Part 4 draws on the recent empirically-grounded work of Sunstein and Dorf to suggest how one ought to go about testing the norms that inform our legal theory, and how such experimentation alters simultaneously both the facts and norms with which we work – and thus the practice and the theory of law.

2 The natural, the practical and the theoretical

I have a magic Uncle Harry. Just last year, he made a little white ball disappear. Standing on the 15th tee at the Riverside Golf Course, approximately 140 metres from the green, Harry hit a solid five iron that covered the flag. It bounced once, twice, and rolled straight into the cup. His playing buddies, Irwin and Jake, shouted: ‘It went in!!!! It went in!!!’

My uncle, the master of prestidigitation, had seen nothing. His failing
eyesight made it impossible for him to watch his ‘lucky dunk’. But the ball
never lies in golf – and when Harry, Irwin and Jake reached the 15th green,
Harry’s ball reappeared – nestled snugly at the bottom of the hole. He has the
plaque on his wall at home to prove it.

Harry is, was, a natural. A schoolyard legend in Brooklyn. When I
picked up the game a few years ago – after a hiatus of 25 years – I did not
have my uncle’s untutored, unstudied magic upon which to draw. I
struggled. I wish that I could say that all that my coach Constanza and I had
to do was fix what was broken. But that would assume that we really had
something with which to work. As it turns out, there is nothing natural or
magical about the golf swing.

Constanza’s first attempt at helping me to create a repeatable,
effective swing failed. My body had a default position – developed from
years of playing baseball, tennis and football – that ran counter to the
requirements of a repeatable golf swing.10 So we started again.

First, we began with my address. We kept my feet planted and still –
throughout the swing. Second, we remained focused on my address. Now I
stood straight and bent slightly at the hips, ‘eyes over the ball’, ass out. Only
the slightest flex of the knees was allowed. For the untrained body, this exercise
was hard: For all the tension (work) lay in my lower back, glutes, hamstrings
and quadriceps. Third, we remained focused on address. My hands would now
hang down, relaxed, before I initiated my takeaway with my left shoulder.
Fourth, we began to work on undoing the flatness of my swing. Here’s a truth
about the body: It will lie to you. It could never lie to Stansi. So while it
seemed to me (on the course) that my backswing was vertical – and though
it even appeared okay on videotape – Stansi could see on the range what I could
not. My shoulders, arms and hands were not where they were supposed to be. So
we worked on my takeaway – a solid feature of my swing. Now, however, my
arms and hands came up and away from my sternum and my wrists cocked at
the top. It felt – and looked – from my perspective on the range – entirely wrong.
But again, the videotape does not lie. They were where they were supposed to
be. We worked still more on my finish: hips and chest turned toward the target,
arms relaxed behind my head, weight posted firmly on my left side

Solution. Success. Not so fast. After many, many months, we had
reprogrammed my body and my brain so as to create a repeatable swing. But
after a year of hard academic work and almost no golf, my defaults returned
with a vengeance. And so we started again – keeping what was still good:
remembering to stay rooted and quiet in my lower body, turning along the
same axis back and through the ball to a proper finish. Constanza, my
teacher, could apply her keen eye and understanding of golf theory to the

10 For more on the problems of default positions (and how one resets them), see
C Sunstein Infotopia (2007); R Thaler & C Sunstein Nudge (2008).
movement of my body. But it remains up to me to make the practice and theory meaningful. The ball never lies. If it goes on an errant course, however, I know enough to ask several questions: Where did I finish? Were my feet rooted? Did I cock my wrists at the top of the swing? Was the tempo of my swing the same going forward as going back? Did I finish with my weight on a straight left leg, arms relaxed, with club and hands behind my back. Instant empirical feedback and an opportunity to put a little theory back into practice.\footnote{Don’t believe me? Listen to Hank Haney – Tiger Woods’s previous coach – explaining what most amateurs do not do. They fail to deconstruct and reconstruct their swing when the flight of the ball tells them that something must be wrong. Here is the question asked of Haney (8 October 2009) by Connel Barrett: ‘What’s the biggest mistake we mortals make in the swing?’ Haney replies: ‘The average golfer doesn’t correctly diagnose the problem in his swing ... Instead they experiment with what might work, or with what they read in a magazine. Or they go after a “feeling” in the swing. What you should do instead is ask, “What’s my golf ball doing? What’s my ball flight?” ... You have to be a detective, and work backwards from impact. Before a lesson or a range session, start with a ball-flight diagnosis and work from there’.}

The first moral of the story: We can all use a good coach – in golf and in virtually all of our endeavours.\footnote{Psychotherapy is another such practice. One initiates therapy because one believes that the manner in which one engages the world is flawed and leads to less-than-optimal outcomes. The therapist provides two goods that the individual alone cannot provide. First, she presents a relationship – a context – in which the patient can act out general patterns of behaviour in the world. Second, she can provide a critical voice that sets out the individual’s maladaptive behaviour in sharp relief. While there is no substitute for the internalisation of this critical voice, we must be able to see the errors first before we can respond constructively to them. Over time and with greater experience of the therapeutic process, the patient begins to note certain regularities in her behaviour and how she and the world respond to her various modes of engagement. As time goes on in therapy, certain kinds of preferred ways of being should become ‘second nature’. That is, in therapy, and over time, we replace existing maladaptive dispositional states with beneficial dispositional states in such a manner that we ultimately require absolutely no conscious awareness of those states.} More importantly for our story, we can even use a good coach in law: We need feedback supplied by coaches and critics who can see what we cannot – where we are going, and where we are going wrong – and we need exercises designed so as to train ourselves not to constantly fall backwards into our unhelpful default positions. Again – what’s true in golf is true in law.\footnote{As remarks made in further footnotes make clear, the extended analysis of such divergent social practices such as therapy, golf, baseball and law goes beyond the notion of ‘practice makes perfect’ or ‘instrumental reasoning’. These practices are practices, not mere analogous activities – and theory and practice in both take largely the same shape. My respondents contend that the practice of golf is only concerned with putting the ball in the hole, and that theory only relates to improving a score. That proposition is fundamentally false. As already explained, golf happens to be extremely norm-governed (around the principle of fairness). More importantly for the purposes of this exchange, it is both a norm-creating and ethically-engaged social practice.}

The second moral of the story: Watch yourself swing a golf club on film: then bring the insights and collective wisdom of millions of golfers – and your professional golf coach Stansi – to bear on what you are doing right and what you are doing wrong. Go out and repeat the drills designed to bring it all together. Suddenly you are part of a social practice in which centuries of trial
and error put us in a position to correct mistakes – at the individual level. Go

down to the pro shop and look at the new range of clubs designed to improve

play. All of a sudden you are part of a social practice in which centuries of trial

and error have put manufacturers of equipment in a position to make

improvements to ball and club design – at the collective level. Golf, so

understood, reveals itself to be a social practice in which the ‘downward

causation’ of multiple participants enables all boats to rise – at least a little –

without any one person or any one authority responsible for such a rise.14

3 Legal theory and legal practice

3.1 Law and the academy

Golf is simply one of many conscious norm-governed and norm-creating

practices that enable us to focus on aspects of our environment, while

simultaneously holding them up for critical scrutiny, in order to form better

responses (descriptively and prescriptively) to the environment.15 Legal

academic life is another such practice. This reply itself is a product of such a

practice. It relies to a significant degree on the legal, philosophical and scientific

contributions of others. (Indeed, the originality of the paper lies primarily in the

application and the synthesis of these other contributions to philosophy, social

theory and legal philosophy.) More importantly for my argument, legal

academics – the legal community’s theorists – function as (one of many)

feedback mechanisms for error correction and truth propagation within South

---


is a practice that dates back several hundred years. Moreover, it is now a practice that

has generated significant amounts of study as to what the body does during the swing

and how the body’s movements can be orchestrated in a manner that produces the

greatest and the most consistent amount of accuracy, power and control.15 That golf is a norm-governed, critical, reflexive and norm-creating practice at a number of

different levels is evident to anyone with even a passing familiarity with this form of life.

We can start with the rules of golf. A deceptively small (but thick) book lies in my study.

Given the disarmingly simple title Rules of golf, produced (and updated regularly online)

by the Royal and Ancient Golf Club of St Andrews (R&A), it tells you, in detail and at a

level of complexity that might defeat the average tax lawyer, the rules that govern the

game of golf. Few other legal regimes apply, and are enforced, globally in a manner that

is so designed to ensure fairness everywhere. Moreover, the rules of golf constitute one of

the few legal regimes in which self-enforcement is the norm. The stakes for professionals

are quite high. On two occasions during the US Professional Tour in 2010 (roughly 35

tournaments), players called penalties on themselves that resulted in the loss of the

tournament. Golf actually requires ethical behaviour of the highest order: Indeed, it is

deeply committed to justice as fairness. Law only requires that one does not get caught:

golf requires that one catch oneself. Moreover, both the USGA and the R&A regularly

alter the rules to ensure that justice as fairness remains the guiding principle of golf. After

changing the rules on the shape of grooves in clubs in 2010, both the USGA and the R&A

took cognisance of the Orwellian role of television in golf and generated a new rule in

terms of which a delayed recognition (post signing of a scorecard by the golfer) of an

infraction (caught by anyone, including the golfer, with access to a live or a subsequent

televised broadcast) results only in the imposition of a two-stroke penalty and not the

heretofore Draconian disqualification of the golfer from the entire tournament. The

purpose: fairness, not retribution.
The relationship between theory and practice across forms of life

Africa’s constitutional order. At a minimum, the legal academy points up logical or empirical flaws. At its best, this knowledge system provides both the grounds for understanding the (legal) world and the conditions for offering new and better ways of being in that world. Finally, law is just one of the many social practices that govern our life shot through with normativity.

3.2 Problems with playing ‘GOTCHA!’: Dworkinian maximalism v Fishian anti-foundationalism

In one of his many tiffs with Ronald Dworkin on the relationship between legal theory and legal practice, Stanley Fish found himself caught up in a side-bar about the relationship between theory and practice in baseball. (And for our purposes baseball will serve quite well as a stand-in for the relationship between theory and practice in golf (and thus law).)

After conceding Dworkin’s point that ‘explanation and performance skills’ are closer in some practices than others – and that indeed judging may be one of those practices – Fish makes an error in playing ‘gotcha’ with Dworkin over the place of ‘theory’ in baseball. It is an error in large part because Fish had already struck out Dworkin by demonstrating that theorising in judgments is largely a rhetorical strategy designed to make the findings and the holdings of a case more compelling. Score it as ‘strikeout + wild pitch’. But it allowed Dworkin to advance to first base when he should have gone back to the bench.

Fish catches out – or so he thinks – Dworkin’s limited understanding of the relationship between baseball theory and baseball practice in the following set of quotations, a dialogue that Fish largely constructs and that I redact and comment upon. I set out the ‘conversation’ so that readers themselves might follow this fight between two heavyweights and the opposing positions that they set out:

16 I have made this connection clear elsewhere:

‘Our mission [as academics] then is to step back and suggest how a body of law hangs together – or if it does not cohere, then to explain why it falls apart. In both instances – the good faith reconstruction, on the one hand, the preferred reading on the other – we offer judges and lawyers the intellectual scaffolding upon which to better – a more just – legal system ... Those of us who work in the guild of academic law-masons – whether academics, jurists, practitioners and students – cannot for a moment sit idly back and assume that the Constitution and all our lovely political institutions provide an answer for the depredations of apartheid or the desolations of the current political order ... Every act in everyday South African life ... possesses an unavoidable moral salience. Or rather, we avoid that moral salience – and the sacrifices those acts entail – at the peril of our souls’.


17 My friends David and Joha accord law a special place in the norm governed worlds we inhabit that it simply does not deserve.

18 S Fish ‘The jurisprudence of Richard Posner, Richard Rorty and Ronald Dworkin’ in There’s no such thing as free speech, and it’s a good thing too! (1994) 200 225 - 230.
(a) Dworkin writes: ‘Denny Martinez never filed an opinion.’\textsuperscript{19} Denny Martinez was a famous pitcher for the Boston Red Sox who denied that any given pitch had a ‘theory’ behind it. Thus, the practice of baseball and the practice of judging – according to Dworkin – could not be more different.

(b) Fish offers this reply: ‘Meaning I take it that Martinez’s resistance to theory [on Dworkin’s view] ... can be explained by the fact that the activity [Martinez] is engaged in is not a reflective one, as opposed say, to the activity of lawyering ...’\textsuperscript{20}

(c) Fish then continues his assault on Dworkin’s position: ‘In Dworkin’s view this distinction marks a hierarchy in which the reflective practitioner is superior, even morally superior, to the practitioner who just goes about his business. Indeed, so committed is Dworkin to this valorisation of the reflective temperament ... that he finds it in places he has just declared it absent.’\textsuperscript{21}

(e) Dworkin contends: ‘Even in baseball ... theory has more to do with practice than Fish acknowledges. The last player who hit .400, fifty years ago, was the greatest hitter of modern times, and he built a theory before every pitch.’\textsuperscript{22}

(f) Fish responds: ‘Now I yield to no one in my admiration for Ted Williams, who has been my idol since boyhood ... But the fact that Williams was a student of hitting and wrote a much admired book on the subject ... does not mean either that his exploits are the product of his analyses or that players ... less inclined to technical ruminations on their art were lesser performers. Williams himself testifies to the absence between theory and practice when he reports the view of another “thinking man’s” hitter, Ty Cobb. “Ty Cobb ... used to say the direction of the stride depended on where the pitch was – inside pitch, you bail out a little; outside you’d move toward the plate”’. Of Cobb’s analysis Williams says flatly, “This is wrong because it is impossible”, and then goes on to explain that given the distance between pitcher and batter, the speed of the ball, “you the batter have already made your stride before you know [consciously] where the ball will be or what it will be”. What this means is that Cobb, possessor of the highest lifetime batting average in baseball history (.367), did not understand – theoretically that is – what he was nevertheless doing better than anyone else who ever played the game ... Later in the book Williams calls Cobb the smartest hitter of all and we must assume that his intelligence ... had nothing to do with his theoretical skills, which were, as Williams has demonstrated, deficient ... What was the content of the intelligence Williams so much admires: ... “thinking it out, learning the situations, knowing your opponent, and most important, knowing yourself”... The point becomes clear when Williams declares that “guessing or anticipating goes hand in hand with proper thinking” ... The smart hitter, in short, pays attention ...’\textsuperscript{23}

\textsuperscript{19} R Dworkin ‘Pragmatism, right answers and true banality’ in M Brint & W Weaver (eds) \textit{Pragmatism in law and society} (1991) 359 368.

\textsuperscript{20} S Fish ‘Almost pragmatism: The jurisprudence of Richard Posner, Richard Rorty and Ronald Dworkin in Fish (n 18 above) 200 228.

\textsuperscript{21} As above.

\textsuperscript{22} Dworkin (n 19 above) 382.

\textsuperscript{23} Fish (n 20 above) 229. Perhaps Professor Fish would do well to watch the Brad Pitt movie \textit{Moneyball} (2011) – or he could read the book – to see how statistics radically altered the manner in which baseball players are assessed and how teams are constructed. The answer: By statistics that rely upon complex algorithms.
(g) Fish concludes: ‘Now it may be that when Dworkin speaks of the theoretical component of practice he is referring to nothing more than the habits of being alert and paying attention ... But somehow I don't think so, for in every formulation that seems to bring us closer together there is something that opens a gap.’

(h) Dworkin: ‘[A good judge] ... will naturally see that he must be, in Fish’s terms, a theoretician as well, and in virtue of, occupying his role as a participant.’

Fish so abhors theory that Dworkin’s last parry is rejected out of hand and to a great extent leaves us no closer to understanding the relationship between legal theory (Dworkin) and legal practice (Fish). But that is Fish’s problem – and a function of his errors regarding the relationship between theory and practice in baseball. What do I mean?

No sport of which I know draws as heavily on statistics as baseball. And I am not just referring as a start to those stats to which the reading public has access.

Over 30 games of spring training, 162 games in the season, almost 20 in the playoffs (210 games for a championship team in a single eight-month season!), a team gathers an almost indigestible amount of information about their own team and, as importantly, the other teams that they play. They learn everything from pitchers’ strong suits, preferences and abilities to batters’ tendencies and weaknesses. This highly accurate data is generated by computers that reflect hot spots and cold spots in the batter’s strike-zone (in the past, players and coaches kept notebooks), where balls are hit, where the location of a pitcher’s pitches are, as well as where hitters like their pitchers to be (more hot spots in red, cold spots in blue). The enormous amount of data is collated and reviewed – film is watched and the

24 Fish (n 20 above) 230.
25 Dworkin (n 19 above) 382.
26 American football relies heavily on stats, but even more heavily on film that is not accessible to the public. Players spend hours watching film in preparation for the week’s game. The 2010 National Football League Championship turned on a play that can be directly traced to the intensive study of film. The Indianapolis Colts, led by Peyton Manning (this generation’s greatest quarterback), demonstrated a tendency, picked up by New Orleans Saints safety, Darren Sharper, to run a specific play when they lined up in a particular formation before the snap of the ball. With Dallas Clark in the slot, and Reggie Wayne inside him, chances were high that Manning would throw the ball to Wayne. Sharper recognised the play taking shape before him as the Colts move to the line of scrimmage late in the final quarter. Manning dropped back to pass. Wayne ran the expected short crossing route. But instead of a complete pass, Sharper anticipated the destination and the timing of Manning’s throw perfectly, stepped in front of Wayne, intercepted the pass and ran the interception back for a touchdown that, effectively, decided the Super Bowl. That’s sports! For the uninitiated, it might be hard to tell the extent to which a theory played a critical role in the outcome. But it was Sharper’s theoretical understanding of how the Indianapolis Colts operated – until that point in time – that enabled him to make the appropriate move when it came time to put theory into practice. Pace Bilchitz and Tuovinen, Sharper’s response was neither non-cognitive nor merely habitual. He had enough time to survey the Colt’s alignment and make a well-informed hypothesis. His hypothesis proved to be true.
data is poured into algorithms that spew out probabilities generally borne out in reality. Remember probabilities are statistical averages: However, if an algorithm supported by the evidence and the probabilities tells you a left-handed batter virtually never hits to the left and opposite side of the field, then you would do well to shift your fielders predominantly to the right side. (The ‘shift’ works – though not all the time.) War games offer a similar opportunity to test a battlefield strategy before a war has begun – to see how it will play out. (Politicians – unlike military officers – often ignore the war game’s tested exit strategy, given that it was not the reason for military action in the first place and (generally) has no effect on their subsequent attempts at re-election.) Baseball coaches who similarly ignore algorithms and continually forget to form a viable exit strategy are invariably fired.

I have, in passing, mentioned the huge amount of film that coaches and players watch in order to form an accurate picture of what they are about to confront. To suggest that the manner in which information is organised and hypothesised about – as Fish does – is not ‘a set of theories’ suggests a vexed philosopher and not an avid baseball fan. Fish is certainly correct in suggesting that Dworkin is rather clueless about baseball. No batter goes up to the plate with a theory of everything, let alone a theory before each pitch. However, batters have spent inordinate amounts of time watching pitchers (pitches are even charted as the game occurs in real time to tell future batters what to expect), and they go up to the plate with pre-dispositions based upon their own strengths and the perceived weaknesses of the pitchers on the other team. A batter might think: ‘He can’t get his curveball in for a strike, I will wait on his fastball’. Or a team might think, ‘his split-finger fast ball’ (a slightly slower pitch with lots of movement) is ‘falling off the table tonight’. (That’s a good thing for the pitcher because it completely fools the batter. The batter ‘sees’ a straight hard pitch aimed around his belt only to have it arrive at the plate more slowly and dive down around his ankles.) The batter has a response, since the split-finger change-up is often not a strike (though it may initially look like it will be one). The batter (and his teammates and coaches) may think – theorise – that since this effective pitch is not going to be a strike, ‘I’ll wait on a fastball’. (Of course, that is extremely hard to do. The pitch is designed to look like a fastball as the pitcher releases it, and as we have shall see in a moment, no time exists in which to make conscious adaptations.) These theories – based on past and immediate experience – the batter takes with him to the plate.

Fish overplayed his hand.

Fish would have been well-served by stopping his argument after catching Dworkin out in the beginning. But, again, he makes a game-losing error by continuing his attack in this game of GOTCHA. Ty Cobb did not understand that the time between the release of the ball by the pitcher and its arrival at home plate (.45 seconds) is faster than the conscious
The relationship between theory and practice across forms of life

Awareness of the pitch (.6 seconds). No batter can employ a grand theory before every pitch. What would such a theory tell you since the action would outpace – in real time – the conscious awareness of the pitch and the swing? Yes: That’s right – the hitter has initiated his swing before he is consciously aware of the fact that he has swung. [Recall Cobb’s error.] However, Fish has made a grand mistake about consciousness and theorising. Baseball players, after 10 000 hours of playing and more, have well-established pre-dispositional states that allow them to react more quickly to novel situations than they are consciously aware. They have built-in theories that allow for extremely subtle adjustments to the pitch headed in their direction.

However, it does not follow from the batter’s habitual – as opposed to conscious – response that ‘theory’ plays no role in baseball. From pitch counting (how tired is a pitcher likely to be after throwing 90 pitches, and thus in less command of his repertoire), from watching his location (is he in control or command of his pitches from the very beginning) to the batter’s own self-awareness, a batter may walk about to the plate with a ‘feel’ – an extremely well-informed feel – for what must be done. Fish is absolutely, categorically wrong about the (lack of) existence of theory in baseball. It exists – in an extremely refined fashion that proves of great use to all of its practitioners.

But of course there is no time for conscious reflection as one initiates a swing. We practise. We may reflect between pitches. We may take lessons from one day and apply them to the next. But we still have to play – and we play at our best after years of practice and ‘theorising’ about what works and what does not. So pace Fish: We spend a career practising and theorising about the game – the two work hand in glove to enable us to do what comes

27 The work of empirical psychologists Benjamin Libet and W Grey Walters has provided a well-established framework for understanding delayed conscious awareness of ‘unconsciously’ initiated action. See, eg, B Libet ‘The experimental evidence for subjective referral of evidence backward in time: Reply to PS Churchland’ (1981) 48 Philosophy of Science 182; B Libet et al ‘Time of conscious intention to act in relation to the onset of other cerebral activities (readiness potential): The unconscious initiation of a freely voluntary act’ (1983) 106 Brain 623; Walters’ Presentation to the Osler Society (1963) as reported in D Dennett Consciousness explained (1991) 167 - 171. Some readers might want to know how would one explain the fact that while a baseball travels the 60 feet and 6 inches from pitcher to hitter in 0.45 seconds, Libet’s experiments seem to reflect a much more generous period of 0.8 seconds between readiness potential and action. The explanation is agent-priming. Constant habituation enables actors to shorten dramatically the period between non-conscious intention and action.


naturally (and here we both mean without immediate conscious deliberation) when one has to bat in a live game.

What unifies this discussion of such different ‘forms of life’ as golf, baseball and the law is the interest that Dworkin, Fish and I have in the relationship between theory and practice in most (if not all) social practices, and what bearing our respective conclusions have for the role of legal theory in legal adjudication. Needless to say, whereas Dworkin appears to contend that theory is everywhere and everything in baseball as in law, Fish does his best to demonstrate that theory – in baseball or law – is of marginal concern because it has such a marginal effect in these social practices. Neither reification of theory nor its rejection truly explains what goes on in virtually all sophisticated social practices. As I have been at pains to point out, theory and practice fit hand in glove in sports such as golf and baseball, as much as they do in ostensibly more reflective endeavours such as legal theory and legal adjudication.

Why should one prefer my view of matters – and reject the conclusions to be found in Dworkin’s and Fish’s game of GOTCHA? On my account, conscious theorising is actually a product of a neurological system and social endowments that have a three-fold purpose: (a) ‘durable and explicit information maintenance’; (b) ‘novel combinations of operations’; and (c) ‘intentional behaviour’. In more common parlance, (a), (b) and (c) describe (1) memory; (2) thought experiments based upon prior experience; and (3) actions designed (subsequent to thought experiments or theoretical analysis) to realise optimal outcomes (both descriptive and prescriptive.) That is, as Dennett argues, our conscious beliefs, about law, baseball or golf, function as ‘idealised fictions’ that enable us to engage – in advance – in a sophisticated ‘action-predicting, action-explaining calculus’.31

That does not mean that theory and practice operate in identical manners in all three social practices. My aim is quite modest: It is to ‘press’ practitioners and judges to take greater cognisance of ‘theory’ in their work by demonstrating just how much work constant theorising does in sports. (If you had watched a pained New York Yankees manager Joe Girardi walk out of the stadium in October 2010 holding two thick binders full of statistics about past performances and future probabilities – having just made an errant set of decisions (based on the books) that may have cost the Yankees the American League Championship – then you might better understand the degree to which theory plays a major role in what many wrongly view as non-cognitive practices.) At the same time, I mean to press the less spectacular proposition on my South African academic colleagues who work constitutional law that they really operate at the periphery of practice and that ‘the illusion’ of theory’s power is primarily a function of the sometimes genuine, and sometimes merely apparent, underdevelopment of this body of law in South Africa. (Advice: Play for

---

citations while you can (and if you have the aptitude), then settle for a therapist’s more indirect affect and consolation.)

4 Whatever works: The case for experimentalism in South African constitutional law

The time has arrived to turn from golf and baseball to the matter at hand. Another set of challenges to Dworkinian maximalism – and its reification of theory – comes from two related quarters: choice architecture and experimental constitutionalism. Both turn on a particular response to and rejection of the kind of grand theorising that Dworkin offers us in works like Law’s empire. For reasons set out below, these two realist responses have far greater traction with respect to South African constitutional law than Fish’s rabid super-pragmatic rejection of theory. There is, however, another challenge to theory on the South African playing field – avoidance, or South African minimalism. It is to an engagement with this anti-theory theory that I turn before identifying ways of building a better mousetrap.

4.1 South African minimalism and avoidance

In its least pernicious form, the core principle of minimalism has been articulated by the Constitutional Court in Mhlungu as follows: ‘[W]here it is possible to decide any case, civil or criminal, without reaching a constitutional issue, that is the course that should be followed.’ On its face, this salutary rule seems unobjectionable. What is, objectionable, even on the Court’s own terms, is turning this salutary rule into a full-blown jurisprudence in which a court must never ‘formulate a rule of constitutional law broader than is required by the precise facts to which it is to be applied’. That rule sounds very much like what Deputy Chief Justice Moseneke has in mind when he writes:

First, those who plead cases before court are themselves steeped in a tradition that seeks to preserve rather than innovate legal reasoning and rules, particularly within the sphere of the common law. The result is that reliance on constitutional provisions is often half-hearted and an afterthought. Sometimes the constitutional points taken are not borne out by the factual matrix. It is trite that courts cannot fabricate points for parties; they have to do with the case materials at hand.

My critique of such South African minimalism takes the following five-fold form. The first objection is that this early statement in Mhlungu flatly contradicts the Court’s later statement in Mhlungu as to the nature of

32 S v Mhlungu 1995 3 SA 867 (CC) para 59.
33 Zantzi v Council of State, Ciskei 1995 4 SA 615 (CC) para 8.
constitutional interpretation. The Constitutional Court in *Mhlungu* avers that constitutional interpretation takes the form of ‘a principled judicial dialogue, in the first place between members of this Court, then between our Court and other courts, the legal profession, law schools, parliament, and, indirectly, with the public at large’. However, if a court refuses to say more than is necessary to decide a case on its facts, then one can hardly expect any meaningfully predictive principle to be drawn from the judgment. That, of course, leads to the second objection. The lack of precision and almost casuistic approach to constitutional norm setting means that it is difficult for any actor – another lower court, a government official or a private actor – to anticipate future forms of law or conduct that would or would not satisfy the basic law’s general norms. If there is no rule of law to which a state actor or private actor knows that she must conform her behaviour, then it would be surprising to find her attempting to conform her behaviour to some unarticulated and inchoate sense of a ‘rule’ that is consistent with the Constitutional Court’s understanding of what the final Constitution permits. Thus, the third objection: The absence of rules of law undermines the ability of other branches of government to comply with the Bill of Rights – and places the Court in the unnecessarily uncomfortable position of having to reject or to accept government’s positions in any it were ruling *ab initio*. A fourth objection is that the absence of clearly articulated rules undermines rational political discourse. Reasoned disagreement can only take place when parties agree on the general terms of the debate. The Constitutional Court must, in terms of its institutional role, establish the meaning of constitutional norms and thus the general framework for contestation. The Constitutional Court abdicates this institutional responsibility to model rational political discourse by refusing to state, in a comprehensive manner, the reasons that ground its conclusions. A fifth objection is that avoidance undermines the ‘integrity’ of the legal system. It is impossible to create a more coherent jurisprudence without identifying the rules – and the reasons – that ground decisions. Without rules, the Court is free to decide the case on the factual merits – without any recourse to the rules generated by substantive provisions of the Bill of Rights. One potential unintended consequence, would be a Bill of Rights jurisprudence that is half-hearted and an afterthought. Though I consciously stop short of saying that is what we indeed possess.
4.2 Choice architecture

As Sunstein himself recognised in his *ur*-text on minimalism – *One case at a time* – this jurisprudential approach (of minimalism, not avoidance) only secures traction because it is parasitic upon a deep and widely shared set of constitutional norms and (tacit) assumptions held by judges, lawyers and citizens that determines most legal and ethical questions, and only leaves difficult, marginal questions undecided. On the necessity, and the presence, of a solid core (in a well-developed body of law), Sunstein writes:

> Anyone who seeks to leave things undecided is likely to accept a wide range of things, and these constitute a ‘core’ of agreement about constitutional essentials. In American constitutional law at the turn of the century, a distinctive set of substantive ideals now form that core.\(^37\)

Sunstein still hews to the propositions above. However, in the decade that followed *One case at a time*, Sunstein’s project moved away from a ground clearing Cartesian skepticism with regard to theory. Instead, he has adopted a consciously experimental approach to social phenomena and to policy decisions in which ‘space’ or ‘choice’ architects draw on actual data in order to ‘nudge’ stakeholders in a ‘form of life’ into making more optimal choices, or generally more *objective* optimal choices than they would have made if not stuck in their existing pre-reflective and untested default positions.\(^38\)

A few examples might help. Markets, though often imperfect, rely upon limited ‘shared’ information (sometimes no more than price) and generate substantially more efficient outcomes than centrally-planned economies. Some open-source software, like Linux, produces incredibly rich results without any central planning. The web itself – the environment for Linux – produces both optimal and suboptimal outcomes, depending on how information is solicited and how further co-operative endeavours are organised. The best example – with which Thaler and Sunstein open up their book *Nudge* – is a ‘bit’ called ‘The cafeteria’:

> A friend of yours, Carolyn, is the director of food services for a large city school system … [H]undreds of thousands of students eat in her cafeterias everyday.


\(^38\) See C Sunstein *Infotopia: How many minds produce knowledge* (n 10 above; Thaler & Sunstein *Nudge* (n 10 above); C Sunstein *Rumours* (2010)). Bilchitz and Tuovinen make two critical mistakes when engaging Sunstein’s work. First, they fail to appreciate that Sunstein’s minimalism requires norms and theories as much as Dworkin’s maximalism does. Second, they fail to recognise that, despite what Sunstein wrote in 1996/1997, their differences are not merely of ‘mood’, but of ‘methodology’. Dworkin is an unabashed rationalist. Sunstein is an empiricist – a social scientist who puts his hypotheses to the test and only draws conclusions after the results are in.
Carolyn has formal training in nutrition (a Master’s degree ...), and ... is a creative type who likes to think about things in nontraditional ways. One evening, over a good bottle of wine, she and her friend Adam, a statistically-oriented management consultant who has worked with supermarket chains, hatched an interesting idea. Without changing any menus, they would run experiments in her schools to determine whether the way the food is displayed and arranged might influence the choices kids make. Carolyn gave the directors of dozens of school cafeterias instructions on how to display the food choices. In some schools the desserts were placed first, in others last, in still others a separate line. In some schools, the French fries, but in others the carrot sticks, were at eye level. From his experience in designing supermarket floor plans, Adam suspected that the results would be dramatic. He was right. Simply by rearranging the cafeteria, Carolyn was able to increase or decrease the consumption of food items by as much as 25 percent. Carolyn learned a big lesson: School children, like adults, can be greatly influenced by small changes in context ... Carolyn is ... a choice architect. A choice architect has the responsibility for organising the context in which people make choices.39

Thinness (or minimalism) is, on the new Sunsteinian account, not a virtue in itself (if it ever truly was). Understanding one’s environment, testing what works and what does not work in that environment, putting the positive results of one’s experiments into practice, being a choice architect – now that is a virtue. Minimalism, in its coarsest form, allows people to fall back into ‘unhealthy’ default positions: continuing to smoke while pregnant because no information – on the packet or elsewhere in public – has suggested that the behaviour could be dangerous and that one could do otherwise. (Eat carrots to satisfy that oral fixation.) The nudging (or libertarian paternalism) Thaler and Sunstein describe does not deny individuals their autonomy. The nudges provide cues that tend to push people in the direction of ‘better’ decisions (from the point of view of the individual and the system as a whole). Like it or not, judges and other political actors are ‘choice architects’: they frame the environment within which we citizens operate. It is, therefore, wrong for any judge to say she is being neutral by being a minimalist. As Thaler and Sunstein note: ‘There are many parallels between choice architecture and more traditional forms of architecture. A crucial parallel is that there is no such thing as a “neutral design.” ... A good building is not merely attractive; it also works.’40 Moreover, being a ‘good choice architect’ is within the reach of judges and other state actors. As I suggest below, courts that share responsibility for constitutional interpretation, attempt to extract as much information from as many stakeholders as possible, and understand that the norms they set are simply rolling best practices that can be tweaked and reconceived over time are the kinds of ‘choice architects’ that South Africa’s constitutional law requires.

Of course, by emphasising rolling best practices and experimentalism, I do not mean to suggest that all of the normative commitments in South

39 Thaler & Sunstein (n 10 above) 1 - 3.
40 Thaler & Sunstein (n 10 above) 3.
Africa’s basic law – or the US Constitution – are up for grabs. (Nor does Sunstein.) Were the South African Constitution to set no clear, normative floor, then we would not have binding constitutional law and our Constitution would be no more than a trifle.\(^{41}\)

For a variety of reasons that have to do with cognitive biases and group dynamics, Sunstein does not think much of the tendency in the legal academy to reify deliberation as a problem-solving mechanism. I share his views about cognitive biases in deliberative group practices.\(^{42}\) But what is of interest to me and import to this article is that he has become something of an empirical constitutionalist when it comes to building better theories about the normative content of a country’s basic law.

### 4.3 Experimentalism and experimental constitutionalism

This last section looks at the work of a growing contingent of constitutional law scholars who have recognised that problems of information deficit, and limited judicial competence can be resolved by a subtle recasting of existing constitutional doctrines, judicial remedies, and special court structures that extract better information and increase the level of acceptable legal norms through the use of *more mindful* interventions.\(^{43}\) Here I set out the theory in brief:

First, experimental constitutionalism relies heavily on co-operation (and shared constitutional interpretation)\(^{44}\) between the co-ordinate branches of government. The general norm-setting practised by the courts is supplemented by various attempts by other state actors and non-state actors at crafting laws

---

41. When reading Bilchitz and Tuovinen’s response, one must keep in mind that both Sunstein and I agree that *deep and widely-shared normative commitments must exist* before social, moral, legal and constitutional experiments secure any traction.

42. See S Woolman et al ‘Evidence of patent thickets in complex biopharmaceutical technologies’ (2012) 53 IDEA: The International Journal of Intellectual Property – (forthcoming). (Several reasons exist for the failure of perfectly ‘rational’ actors to fail to deliver well-designed intellectual property to downstream markets, including the cognitive biases of licence holders and the attributive biases of the participants. In sum, human beings generally tend, just like ordinary owners of upstream biomedical research patents, to overvalue their own contributions and property).


44. Shared constitutional interpretation stands for four basic propositions. First, it supplants the ‘arid’ notion of judicial supremacy with respect to constitutional interpretation. *All branches of government have a relatively equal stake in giving our basic law content.* Second, it draws attention to a shift in the status of court-driven constitutional doctrine. While courts retain the power to determine the content of any given provision, a commitment to shared constitutional interpretation means that a court’s reading of the constitutional text is not meant to exhaust all possible readings. To the extent that a court consciously limits the reach of its holding regarding the meaning of a given provision, the rest of the judgment should read as an invitation to the coordinate branches or other organs of state to come up with their own alternative, but ultimately consistent, gloss on the text. Third, shared constitutional competence married to a somewhat open-ended or provisional understanding of the content of the basic law is meant to increase the opportunities to see
and policies designed to fulfil those constitutional norms. Over time, courts, state actors and non-state actors will have the opportunity to determine whether various ‘socio-political experiments’ have achieved the ends set for us by our basic law (as interpreted by the courts, the legislature, the executive and non-state actors). We will, in instances of policy failure, also have an opportunity to decide whether the norms or the ends set by the courts, the legislature, the executive and non-state actors constitute the best possible gloss on our basic law. Shared constitutional interpretation relies, less upon a principled dialogue about norm setting and more on rough and tumble engagement) between the courts, the co-ordinate branches of government, and the public that is reflexive, as well as forward and lateral looking.

Second, experimental constitutionalism relies heavily on participatory bubbles.45

These bubbles are court-initiated or political structures designed to ensure that ‘meaningful engagement’ about optimal outcomes takes place. For example, a court might set out a general normative framework within which how different doctrines operate in practice and maintain the space necessary to make revision of constitutional doctrines possible in light of new experience and novel demands. In this regard, the Constitutional Court might be understood to engage in norm-setting behaviour that provides guidance to other state actors without foreclosing the possibility of other effective safeguards for rights or other useful methods for their realisation. Fourth, a commitment to shared interpretations ratchets down the conflict between co-ordinate branches and levels of government. Instead of an arid commitment to separation of powers and empty rhetorical flourishes about courts engaging in legal interpretation not politics (Ex parte Chairperson of the Constitutional Assembly: In re Certification of the Constitution of the Republic of South Africa, 1996 1996 4 SA 744 (CC) paras 26 - 28), courts are freed of the burden of having to provide a theory of everything and can set about articulating a general framework within which different understandings of the basic text can co-exist. Indeed, the courts and all other actors have more to gain from seeing how variations on a given constitutional norm work themselves out in practice. Put slightly differently, shared constitutional interpretation creates the space for different actors in different places or with different briefs to try doing things differently, but constitutionally. The different means may show us which ways of doing things that are more (or less) successful. Or the different means may shed new light on – or change our understanding of – the very constitutional ends the varying strategies seek to promote. See MC Dorf 'The domain of reflexive law' (2003) 103 Columba LR 384.

Whereas a scheme of shared constitutional interpretation introduces an experimentalist element into the upper tiers of government institutions, the physical metaphor of bubbles is meant to convey three qualities of experimentation and engagement in smaller-scale institutional processes. First, processes of deliberation are a natural part of ongoing social interactions. They originate when challenges to a given institutional authority accumulate and finally come to a boil: just as bubbles form after pressure builds up and escape to the surface of a liquid. Second, bubbles are meant to suggest limits on the scope of deliberation. Bubbles only enclose a small amount of space – both in terms of the issues debated and the number of participants. Third, bubbles are ephemeral. After satisfactory resolutions emerge from processes of participatory deliberation, the raison d'être for such processes ceases to exist. Participants can return to their more routine lives. Fourth, experimentalists recognise that deliberation does not, necessarily, lead to better outcomes. It can, quite often, lead to a greater polarisation of positions. See S Sturm 'The promise of participation' (1993) 78 Iowa LR 981; C Sunstein Infotopia (2007). Both S Sturm and C Sunstein note how deliberation can lead, over time, to greater discord and less accurate decisions – for a variety of reasons: from the ‘big dog’ problem to a propensity toward ‘group-think’. Even deliberative democrats recognise such limits without sensing a need to forsake deliberation. See A Gutmann and D Thompson Democracy and disagreement (1996).
The relationship between theory and practice across forms of life

The relationship between theory and practice across forms of life shall take place. Two features of these bubbles deserve closer attention. They may not be limited to the initial parties to the ‘litigation’. Other interested stakeholders – amici et al – may participate in the conversation. The aim, again, is two-fold: greater elicitation of information; greater normative legitimacy of any decision ultimately taken. The other feature of these participatory bubbles is that they may not remain within the domain of the courts. We can easily imagine – and have witnessed in South Africa – greater community participation in hearings called by the South African Human Rights Commission or other chapter 9 institutions, national or provincial legislatures, school governing bodies, or other social and political fora. The Constitutional Court has shown itself alive to the need for ‘participatory bubbles’ when provincial legislatures take decisions that affect the lives of the denizens within their boundaries. South Africa, despite the limits imposed by its largely one party state, has the tools available to make participatory bubbles the norm in norm-setting environments.

Third, participatory bubbles lose their cohesion – and the pressure to produce better than zero-sum outcomes – if the courts fail to articulate the norms within which a preferred solution is meant to occur. If experimental constitutionalism is judged to be an attractive set of principles by which to establish constitutional norms (by widespread public agreement) and to assess best practices (by inviting as many stakeholders as possible to design an optimal remedy for a specific social problem) then the jurisprudence of South African minimalism – or avoidance in the South African vernacular – must be one of the first judicial doctrines to go. Several of the Constitutional Court’s recent judgments must leave one concerned about the ability of the Court to articulate such norms. Put somewhat differently, the process of general norm-setting by the courts that initiates a process of rolling best practices by other parts of the state never gains sufficient traction when constitutional norms remain radically under-theorised.

47 See, eg, Shilubana v Nwamatwa 2009 2 SA 66 (CC), 2008 9 BCLR 914 (CC).
50 See M Bishop ‘Vampire or prince? The listening constitution and Merafong Demarcation Forum v President of the Republic of South Africa’ (2009) 2 Constitutional Court Review 318.
However, let us assume that the courts and other political institutions adopted an experimentalist approach—replete with participatory bubbles and a commitment to shared constitutional interpretation. One must remember that one of the virtues of the experimentalist approach is that it is not static. New challenges to the general norms set by the courts, the legislature, and the executive may arise. These new challenges will likely bring a host of new stakeholders—and thus new participatory bubbles. Experimentalism—especially in law and the social sciences—is a compelling answer to minimalism and absolutism because it recognises (a) the ability to extract information to create norms that reflect the lived existence of people affected by the basic law; and (b) that the circumstances governed by the basic law may change in ways unanticipated by the previous constitutional norms set by political institutions (and their stakeholders). Here is the takeaway: The norms created by the courts or understood by some other political body are understood to be rolling norms that will be subject to change where and when the exigencies of the moment require such change.53

This kind of language is not new to South African politics and the basic law that both amplifies and constrains that politics. In one of the first white papers on primary and secondary education in a post-apartheid South Africa, then Minister Bengu wrote:

Policies are stated in general terms and cannot provide for all situations. Our legacy of injustice and mistrust continuously throws up problems which need the wisdom of Solomon to settle. In this protracted transitional period, in which new policies are being developed and implemented, the chances are that we shall collectively make many mistakes, either in conception or execution. They must be recognised and corrected. The possibility of damage will be reduced if new policies are based on knowledge of our charter of

53 As everyone knows from American police television shows, *Miranda v Arizona* established a clear normative basis for the right to remain silent once a person finds herself in police custody. Dorf and Friedman show that the *Miranda* Court openly invited Congress to address those concerns afresh. Congress did not (genuinely) do so. When the government, in *US v Dickerson*, sought to uphold its post-*Miranda* legislation, the *Dickerson* Court noted that too much time had passed (34 years), and that the Court’s normative position in *Miranda* had become accepted by law enforcement officials and citizens alike as normatively legitimate. Dorf and Sabel’s engagement with drug treatment courts does not deny the normative content of the law those courts enforce. Rather, they demonstrate that a court designed to engage particular kinds of problems and provided with information from (and the experience of) multiple stakeholders (counsellors, doctors, treatment facility personnel, lawyers for all concerned parties, police, and the judge herself) have a greater likelihood to overcome various cognitive deficits and arrive at a solution better for the addict (an opportunity for rehabilitation) as well as the commonweal (less recidivism and a less expensive resolution). Experimental constitutionalism recognises that the norms themselves may change over time as participants in a form of life—say drug treatment courts—recognise the kind of interventions that work and the kinds of interventions that do not. The alteration of norms (more or less incarceration, more or less rehabilitation) will flow from the experience of various actors who have worked in this domain over time. Finally, experimental constitutionalism does not eschew a normative framework—nor even a deep, normative framework. Indeed, the very word ‘constitutionalism’ places experimental constitutionalists within a very explicit, specific, and constraining, Western value order—one that South Africa shares (in theory).
fundamental rights and on sufficient consultation with those who are affected by them, if compromises are negotiated, and if principled compromises are sought.54

Minister Bengu’s approach to the problems of South African primary and secondary education in a post-apartheid dispensation fits this notion of experimental constitutionalism perfectly. The Minister understood the basic norms that governed South Africa. He had consulted all of the relevant stakeholders – simply too numerous to mention here. And he, and his national Department of Education, had then generated a white paper that would become the South African Schools Act. At the same time, Minister Bengu recognised that the law and policy reflected in South Africa were always going to be provisional. They represented a good faith effort to solve the problems that faced primary and secondary schools in post-apartheid South Africa. But he could not say, in advance, which policies would serve the general commonweal and whether the norms that underwrite the new social contract in our (then) new commonweal would not, themselves, change over time.

5 Conclusion

Tiger Woods did not become the greatest golfer to ever grace a course by resting on his laurels or sticking with the swing that allowed him to reach the pinnacle of his sport. Indeed, at a point in time when no one would have argued that he was – by far – the best in the game (2005), he changed his coach (Butch Harmon) and began a long process of reconstructing his swing. The immediate result was a patch of play that was erratic – which, in turn, led to a long stretch of disappointing results (read: absence of wins). However, within two years of painstaking effort – theorising and practice, hand in hand, for thousands of hours – Tiger Woods had, with the careful assistance of another brilliant pair of eyes (Hank Haney’s), rebuilt his swing into something even better than that swing which had already smashed innumerable records. He is now repeating the same excruciating endeavour under the tutelage of Sean Foley.

The same sort of commitment to experimentation must drive the jurisprudence of our Constitutional Court. It is not enough to shift the blame, as the Deputy Chief Justice does, when he writes that

... those who plead cases before court are themselves steeped in a tradition that seeks to preserve rather than innovate legal reasoning and rules, particularly within the sphere of the common law. The result is that reliance on constitutional provisions is often half-hearted and an afterthought.55


55 Moseneke (n 34 above) 11.
He is certainly correct that no group of 11 women and men are in a position to lay down basic norms that will resolve all of the potential problems, in a given area, that may seize the Court. But no one has suggested that they need a ‘theory of everything’ that would both dispose of the matter before the Court and all future cases. The only claim pressed in these pages is that if the Court should choose to follow an experimentalist mode of adjudication, then its decisions are likely to be better reasoned (empirically) and more compelling (normatively). Moreover, the Court need not be stuck in time. Its general norm setting should allow for an assessment of rolling best practices, and a rejigging of the general norms that the times and the environment dictate.

Academics have a critical role to play in this experimentalist process of constitutional theory building. We are, first, another set of eyes. We can see – on occasion – that which the Court and its members cannot: Experience, reflection, theories about how different areas of the law work make academics an invaluable resource for the Court. (We may know a good property ‘swing’ when we see one: and we should – if we are André van der Walt or Theunis Roux be able to identify problems and flaws.) We are, second, free from various encumbrances that may limit the Court’s capacity for reflection and theory: Neither the pressures of time, nor the exigencies of the parties before the Court, nor the need for intra-institutional collegiality nor inter-institutional comity detain us. We have only one goal: to read the Court’s judgment’s carefully, to reflect back to the Court what we see and hear, and to make the Court’s future judgments better. We are here to assist the Court in building its various theories – keeping in mind all the time, that what works for one period of time might

---

56 I actually contend elsewhere that the use of ‘meaningful engagement’ in a series of cases over the last few years holds out some hope for actual movement in the direction of experimental constitutionalism. See S Woolman The selfless constitution (n 43 above). See, eg, Head of Department, Mpumalanga Department of Education v Hoërskool Ermelo 2010 2 SA 415 (CC); Residents of Joe Slovo Community, Western Cape v Thubelisha Homes [2009] ZACC 16, 2010 3 SA 454 (CC), 2009 9 BCLR 847 (CC); Abahlali Basemjondolo Movement SA v Premier of the Province of KwaZulu-Natal [2009] ZACC 31, 2010 2 BCLR 99 (CC); Occupiers of 51 Olivia Road, Berea Township and 197 Main Street Johannesburg v City of Johannesburg[2008] ZACC 1, 2008 3 SA 208 (CC), 2008 5 BCLR 475 (CC). The problem with these cases, and others, is that although the Court creates the space for polycentric, multi-stakeholder problem-solving, its persistant refusal to set clearly articulated norms (and its general preference for mere ratification of a settlement agreement) closes down the possibility for genuine experimentation (over time) in other (future) cases and in other (institutional) settings. The various stakeholders in novel cases (or settings) will lack a general set of norms by which their behaviour, expectations and strategies can be assessed, compared, measured and altered. For example, the absence of clear guidelines regarding the content of the right to a basic education in sec 29(1) of the Constitution actually made ‘meaningful engagement’ – as the Court requested in Governing Body of the Juma Musjcid Primary School (Court Order) – impossible and forced the Court to show its cards in Governing Body of the Juma Musjcid Primary School v Ahmed Asruff Essay CCT 29/10 [2011] ZACC 13.

57 AJ van der Walt Constitutional property law (2005).

need to be revised over time, as circumstances, and the environment in which the Court is located, change.
1 Introduction

We face a revolt from theory, in law and across the rest of the intellectual landscape.\footnote{1}{R Dworkin \textit{Justice in robes} (2006) 72.}

The role of theory in law is a topic that has generated much heated debate, particularly in the United States. In South Africa, this debate has largely taken the form of academic exchanges concerning the virtues of judicial minimalism: This involves the idea that in adjudicating cases, judges should confine themselves to providing reasons that are narrowly tailored to the facts before them and are not to go beyond what is necessary to justify that decision.\footnote{2}{Of course, what is necessary to justify a particular decision is itself contentious and will be the subject of discussion later in this piece. Currie leaves this point undecided, stating that he \textquote{has not had an opportunity to explore in any depth the pragmatic and political virtues of minimalism'} \textquote{Judicious avoidance'} (1999) \textit{SAJHR} 138 165; see also C Roederer \textquote{Judicious engagement: Theory, attitude and community'} (1999) \textit{SAJHR} 486.}

This debate has taken place in the context of criticisms that have been lodged against the Constitutional Court for avoiding giving content to fundamental rights and its failure to provide more detailed and deeper justifications for its decisions.\footnote{3}{See, eg, A Cockrell \textquote{Rainbow jurisprudence'} (1996) \textit{SAJHR} 13; D Bilchitz \textit{Poverty and fundamental rights} (2007); S Woolman \textquote{The amazing, vanishing bill of rights'} (2007) \textit{SALJ} 762.}

In the article titled ‘On the common saying “What’s true in golf is true in law”: The relationship between theory and practice across forms of life’, Woolman attempts to capture the relationship between theory and the practice...
Chapter 16

of adjudication. The first part of the article discusses the relationship between theory and practice in golf. The important conclusion of this discussion is reached when Woolman asserts ‘what is true for golf is true for the law’, as both are social enterprises that require feedback, ‘coaching’ and improvement through practice. Woolman then proceeds to engage with the debate concerning theory in law between Ronald Dworkin and Stanley Fish where again a sporting analogy surfaces (between law and baseball this time). The upshot of this discussion appears to be that even in as physical and applied a sport as baseball, theory and practice are intertwined: ‘Theory and practice fit hand in glove in sports such as golf and baseball as much as they do in endeavors such as legal theory and legal adjudication.’ The claim is not that, whilst actively engaged in ‘playing the game’ one is consciously aware of theory, but rather that the hours of practice and theorising create dispositional states that enable a player to improve their performance. The same is true with law: According to Woolman, theory can help condition the dispositions that judges have in deciding cases. It is thus indispensable in helping to improve the judgments that result though it should not be reified either.

Woolman then moves on to engage with judicial practice in South African law. He makes several arguments against judicial minimalism, arguing that even a key proponent of minimalism such as Sunstein appears in recent work to have had doubts about its usefulness. Judges are ultimately “choice architects”: They frame the environment within which we citizens operate. As such, they cannot be neutral in what they do. ‘Understanding one’s environment, testing what works and what doesn’t work in that environment, putting positive results of one’s experiments into practice, being a choice architect – now that is a virtue.’

In light of this message, Woolman believes that the appropriate approach for judges to adopt is an experimentalist one. This involves a dialogue between different organs of state, participation of affected parties and a constant openness to changing norms in light of different circumstances. The key guiding principle for constitutional norms is the idea of understanding the context and testing what works or does not work within that context as well as an openness to changing norms ‘when the exigencies of the moment require such change’.

---

4 The original piece to which this responds has undergone several revisions. We have sought to adapt our response to the extent possible to the latest draft we have received.
6 Woolman (n 5 above) 354.
7 Woolman (n 5 above) 358.
8 Woolman (n 5 above) 358.
9 Woolman (n 5 above) 362.
In conclusion, Woolman seeks also to capture the role of academics in constitutional democracy and their relationship to judges and adjudication. Academics, Woolman contends, provide experience, expertise and reflection in relation to how the law should work: they are resources for the court to use. Moreover, they are free from some of the difficulties faced by the Court and only have one goal: ‘to read the Court’s judgments carefully, to reflect back to the Court what we see and hear, and to make the Court’s future judgments better’. 10

Woolman ends his article by making the point that academics can, through theoretical reflection, help judges to improve their adjudication.

As has been clear from this outline, Woolman’s article engages a wide range of issues, each of which requires extensive analysis. This reply will focus only upon certain of his core claims. We are in a large measure of agreement with Woolman about the importance of theory in law and its relation to practice; yet, we are of the view that there are several respects in which his conception of the role of theory in law is ‘under-theorised’. This reply can thus in some sense be seen as an elaboration of certain of the key claims in Woolman’s article though we will indicate where we disagree. First, we shall consider the way in which the notion of theory is used in Woolman’s piece. This will involve distinguishing between different notions of theory. As we do so, it will become evident that the role of theory in golf and law is not equivalent. In searching for a more adequate account, we are led to engage with Dworkin’s notion of ‘justificatory ascent’ which concerns the relationship between theory and the practice of adjudication. Two problems with his account lead us to prefer a Rawlsian understanding of this relationship based upon the notion of ‘reflective equilibirum’.

Secondly, given our preferred account of the relationship between theory in law and the practice of adjudication, we support Woolman’s critique of minimalism as an approach to adjudication. We seek to build on this criticism by providing a concrete example of how the case law of the South African Constitutional Court in the area relating to the right of access to courts demonstrates some of the pitfalls of minimalism. Thirdly, we consider the tension between Woolman’s advocacy of the importance of theory and his defence of experimentalism. In particular, it is argued that the pragmatist and empiricist dimensions of experimentalism as articulated by Woolman fail to capture one of the central elements of law, namely, its being a normative enterprise. This entails that experimentalism is fundamentally incomplete as a methodology and an account of the relationship between theory and the practice of adjudication. Finally, we shall argue that the account of the relationship between legal academics and the judiciary offered by Woolman can be improved upon. We shall contend that the relationship should be conceived as

10 Woolman (n 5 above) 365.
a mutually reinforcing one with each playing a crucial role in achieving a
desirable reflective equilibrium between the demands of theory and the
exigencies of adjudication.

Part 1: The many faces of theory

At the centre of Woolman’s article is an analogy between the role of theory
in golf and in law. The analogy is intriguing and it is worth questioning
whether it is in fact apposite. Our approach to doing so will involve first
engaging with the question as to what in fact is ‘theory’. Through clarifying
this idea, it is hoped that it will be possible to capture more adequately the
role of theory in law.

Theory is a word that can have a number of senses. To clarify its usage
in the context of this paper, it is useful to begin with those senses which may
be less apposite. The first relevant meaning of theory is the notion of a
‘conceptual scheme’ being involved in the performance of a particular
action. This notion does not involve the idea that the individual in question
is necessarily self-aware of that conceptual scheme. In this sense, all human
activity – whether it be golf or law – is in some sense laden with theory. This
notion has been important in the history of philosophy where it has been
recognised that even the seemingly simple act of perceiving the world
through sense experience involves the application of a conceptual scheme
to the world.11 Some forms of activity engage more complex cognitive
processes and schemes. In his article, Woolman discusses his uncle’s
intuitive ability to play golf. In hitting the ball, the uncle is engaging several
mental processes that are consciously involved in taking a swing. We can
refer to this as ‘thinking in’ a practice. All conscious animals, to a greater or
lesser degree, will be engaged with ‘thinking in’ practices they perform and
applying conceptual schemes to them. Thus, we agree with Woolman that
the practice of golf is in some sense ‘cognitive’. This sense of theory is itself
unhelpful (and possibly trivial) in the current context as it does not assist us to
capture the particular role of theory in adjudication as opposed to its role in
such basic activities as sensory perception of the world.

The second sense of theory that appears more relevant is not simply the
idea of conscious mental activity which involves the application of conceptual
schemes, but the idea of self-conscious conceptual schemes. This idea can be
captured by the notion of ‘thinking about’ and involves centrally the idea of
reflecting about our thought processes or practices and developing
conceptual schemes from that reflection. Woolman seems centrally
concerned with this idea of theory as a ‘conscious critical practice’, and
in his quote referring to Haney’s concern that ‘the average golfer doesn’t

11 This is a key insight of I Kant A critique of pure reason (1781) and has been the subject of
much philosophical reflection in current philosophy. See, eg, J Dancy (ed) Perceptual
knowledge (1988).
correctly diagnose the problem in his swing'. Reflection about one’s swing could help a player recognise the problems with it and, through conscious practice, ultimately help improve the swing. Initially, golfers may be led to certain ad hoc insights concerning the difficulties with certain bad swings and the characteristics of good swings. Over time, such reflection could lead to certain general principles being formed as to what constitutes a decent swing. The systematic development of general principles from the experience of golfers could lead to the development of a ‘theory of the good swing’. This in turn could help those who are learning to play golf to develop a decent swing.

Having understood the notion of how a theory could be developed in golf, the question arises whether the role of theory (in the sense indicated) is the same as its role in adjudicating legal cases. Whilst there may appear to be a surface similarity that Woolman has highlighted in his article, our view is that the relationship on deeper analysis is distinct. The key difference in our view is that, in golf, the role of theory (in the sense of ‘thinking about’) appears to be largely instrumental. Ultimately, the goal is to play the game in the best way possible: This largely involves ensuring that the golfer who stands at a set distance from holes in the ground succeeds in guiding small white balls into those holes in the least number of attempts. Since the practice is established, there is very little to reflect about in relation to the rules themselves, although some small revisions may be made from time to time. ‘A theory of the good swing’, if it exists, would largely be designed around assisting participants in this practice to improve their skill at playing the game which is defined in terms of a set of generally pre-existing rules.

The role of reflective theory in adjudication is different. Its purpose is not simply to improve a pre-existing practice: In large measure, theory – and in particular normative reasoning – constitutes the very practice itself and is intrinsic to its existence. Law is theory-laden in a manner that golf simply is not. We shall provide four arguments to support this claim. First, of crucial importance is the recognition that law is thoroughly normative in a manner that golf is not. Law imposes certain obligations upon individuals and adjudication would have certain real consequences for parties before a court. For individuals to recognise the legitimacy of such obligations and outcomes, there must be a justification for why they are imposed. That justification would immediately require some theoretical underpinning. If the law is imposed by a judge and lacks any justification, then this would amount to coercion with no grounding. Hart has famously described a
situation where a gunman requires a person to obey him on pain of death. In such a circumstance, if the individual obeys it is not out of any sense of being obliged to do so but rather simply out of fear. Without justification for a decision, a judgment would simply amount to coercion and be akin to the gunman situation ‘writ large’. It would lack a sense of why individuals should seek to obey a judgment. Theory in adjudication (of some kind) is thus crucial to the very legitimacy of the practice itself, a question that is not at issue in the field of golf. This is largely because golf is a sport amongst willing participants whereas law has a coercive quality.

Secondly, the ends of law require normative theoretical justification in a way they do not in relation to golf. What outcomes should judges seek to achieve when performing the task of adjudication? There could of course be disputes as to the very ends they should seek to achieve, whether this be justice, stability or something else. Even if we were to agree that they should aim to achieve justice between the parties, the question will arise as to what the content of this notion is: Does it involve some idea of fairness, or is it simply about achieving the greatest happiness for the greatest number? Judges when adjudicating will thus need to have some idea of what they are seeking to achieve and what justice, for instance, requires of them. That will inevitably require some theory. This point outlines another sense in which adjudication is fundamentally normative. Judges must have a sense of what the practice they are involved in aims to achieve and this involves, at least partially, deciding on questions of how the world ‘ought’ to be. In South Africa (and similar systems that place normative concepts at the foundation of the constitutional order), given that section 39(2) of the Constitution obliges courts and other judicial bodies, when interpreting any legislation and when developing the common law and customary law, to ‘promote the spirit, purport and objectives of the Bill of Rights’, it ought no longer to be controversial that adjudication in all spheres of law requires a certain level of ethical engagement. Golf too requires some engagement with the ends of the practice when rules are in dispute though this occurs in a much more limited sphere and is not necessarily involved with ethical questions. Whilst fairness is of course important within any game, its role is ultimately to ensure that the practice of playing the game is not undermined. Fairness is determined with reference to the point of playing a particular game such as

---

15 Hart (n 14 above) 18 - 19. We have applied the example in a different manner to that which Hart originally intended.
16 J Bentham An introduction to the principles of morals and legislation (1781) famously argued that law should aim at justice which is described by the utilitarian principle. J Rawls’s Theory of justice (1973) argues that a concept of fairness is central to justice.
17 Some may claim that here we are assuming a natural law account of the relationship between justice and morality. Yet, most positivists would accept that law – even though it is conceptually distinct from morality and can be identified independently – should seek to achieve moral ends. To this extent, having some conceptual understanding of the demands of political morality is key. To the extent that judges have choices in adjudication (and Hart would recognise that discretion at least exists in penumbral cases), they would also be required to engage their normative understanding of what law should be.
18 This would be so even on a positivist account of the Constitution for the basic law itself directs a judge to ethical reflection.
golf: Indeed, it cannot be said that the very goal of playing golf is achieving normative ends such as ‘fairness’. Moreover, practices such as golf may actually need to draw on concepts of fairness that have been articulated in disciplines such as law. For, in law, normative concepts such as fairness, justice and stability are the very ends of the practice itself whose content must be developed in the process of adjudication. Consequently, law is more thoroughly suffused with normative theoretical elements than practices such as golf since these ethical concepts ultimately are a fundamental goal of the practice itself.

Thirdly, law is fundamentally an enterprise that takes place through the use of language. Practices centrally involving language are different to those that do not involve language. We may reflect upon such practices as eating, drinking and golf in language yet whilst we are engaging in the practice, language is not centrally involved. This means that, to employ the distinction developed earlier, we need not ‘think about’ the practice whilst we are ‘thinking in’ the practice. We need not necessarily reflect about these practices at all to perform them (though such reflection may help us). Indeed, Woolman recognises that ‘thinking about’ golf or baseball whilst playing the game may in fact hinder one’s performance.\(^{19}\) It is difficult to see how this could be the case in relation to a practice such as legal adjudication. Practices that take place in the realm of language such as interpreting literature and law necessarily seem to involve a higher level of abstract thinking. Moreover, it appears difficult to understand how one could engage in the practice of adjudication without ‘thinking about’ what one is doing. To avoid doing so, would involve failing to perform the practice adequately and appear to involve deciding arbitrarily. Adjudication thus necessarily involves conscious, abstract, and reflective thinking.

Finally, the particular practice of adjudication would usually involve the application and interpretation of past precedents, statutes and the Constitution. It also fundamentally involves the application of rules and principles. All these elements involve questions of interpretation which would at least require judges to have some idea of the approach they adopt in this regard. Furthermore, the decision as to which principles or rules apply and their application to particular facts will inevitably involve reflection and the development of some form of theory. Theory is thus irremediably bolted into the very structure of law itself in a way that is not true for golf. In Dworkin’s words, the law is ‘drenched in theory’.\(^{20}\)

Opponents of ‘theory in law’ would claim that they are concerned with a distinct notion of theory from what we have articulated above. These critics are preoccupied with the idea that adjudication does not require the development of grand philosophical schemes in order properly to justify the decisions that result. The concern particularly seems to be about judges developing highly

---

\(^{19}\) Woolman (n 5 above) 353 - 354.  
\(^{20}\) R Dworkin ‘In praise of theory’ (1997) *Arizona State LJ* 353. The article, originally given as a lecture, is also reproduced as ch 2 of *Justice in robes* (n 1 above).
abstract theories in their judgments. Prominent critics of this view include Posner who is against the development of such overarching theories: law in his view is essentially a practical enterprise aimed at making the best decision in particular circumstances and requires rather an orientation towards ensuring the best consequences overall for a decision.\textsuperscript{21} Here it is quite clear that Posner assumes a particular theory as to how judges should go about deciding cases (a point he admits).\textsuperscript{22} The claim, however, by such critics is not that theory can be eliminated from law but rather that it should be minimised and kept at the ‘lowest’ level of abstraction.\textsuperscript{23} The problem with this view is that it is difficult to see why theory should be confined to the ‘lowest’ level of abstraction and what this prescription actually means. Indeed, failing to engage abstractly with theory may simply mean that judges are less aware of the shortcomings of any particular approach they adopt. Higher levels of abstract thinking may in fact improve the performance of judges in relation to concrete cases. It is thus unclear how one can defend a practice that generally seeks to avoid further abstract philosophical reflection, even where this may improve the the quality of decisions.

This criticism requires us to recognise different types of theories that exist in relation to the law.\textsuperscript{24} First, there are theories about the law, for example, theories about what the nature of law consists in, such as positivism, and natural law. These theories aim to explain the nature of law and may deepen a judge’s understanding of what she is doing. They are not though necessarily required for reaching a particular result in a concrete case. Secondly, there are a whole host of moral, political, social, metaphysical and epistemological theories that underlie and shape the law. These theories are ‘embedded’ in the law\textsuperscript{25} and they give the legal system its shape and content. This group contains both grand over-arching theories of, for example, the purpose and functioning of a state as well as more specific theories about more particular issues such as the justification for the existence of a particular right. Finally, theory also involves the notion of legal doctrine where academics or judges, for instance, draw out general principles from specific cases. Doctrine can involve specific theoretical underpinnings of principles and/or even a collection of those very principles that systematise a particular area of the law.

The law itself thus is connected to theory in different ways and at different levels of abstraction. This idea may be captured by recognising different

\begin{itemize}
  \item Posner has advanced his judicial philosophy in a number of monographs, most recently and comprehensively in \textit{How judges think} (2008).
  \item See Posner (n 13 above) 377.
  \item A similar point is made in C Sunstein ‘From theory to practice’ (1997) \textit{Arizona State LJ} 389 391 - 395.
  \item We do not deny there may be other forms of theory, yet, we attempt to categorise the forms of theory most in evidence in the law.
  \item Dworkin (n 20 above) 355 - 357.
\end{itemize}
layers of theory that are nascent within the law. Dworkin attempts to capture this idea through the notion of ‘justificatory ascent’. To understand this idea, one can usefully begin by considering adjudication by his fictional superhero judge Hercules. Hercules would essentially have a grand theory of everything and could deduce the solution to each particular case from the most abstract of philosophical principles down to the finest detail. As mere mortal lawyers with limited capacities, Dworkin recognises that the rest of us must usually approach legal decision making the other way round, moving from concrete cases towards more abstract principles in a process Dworkin terms ‘justificatory ascent’. Justificatory ascent involves the idea that the concrete principles in the area of law with which one is concerned may not answer the particular question before a judge. To provide an adequate response, the judge may need to engage other areas of law. In some cases, this will involve understanding and engaging with political philosophy concerning the most desirable approach to achieve justice in a particular case. From a particular concrete case, a judge may be pushed to higher levels of abstract theory in order to provide a compelling justification to decide the case in a particular way. Contrary to the claims of Posner, for instance, Dworkin argues that no one can determine in advance the degree of ascent that is necessary for any particular decision and this will be dependent upon the context and the particular case at hand. In summary, he argues that ‘legal reasoning presupposes a vast domain of justification including very abstract principles of political morality, that we tend to take that structure as much for granted as the engineer takes most of what he knows for granted, but that we might be forced to reexamine some part of that structure from time to time, though we can never be sure, in advance, when and how’.

Justificatory ascent does appear to capture in part some of our intuitions as to how theory enters adjudication. Yet, we wish to highlight two particular problems with it. The first involves showing why the ascent to a Herculean level of abstraction is not necessary in every case. In other words, every legal case will involve some consideration of applicable legal provisions. This will involve at least some considerations of language as well as certain normative considerations of justice. As such, every case would arguably invoke at least potentially complex questions of language and justice. It is unclear the basis upon which Dworkin claims judges may justifiably ascend to higher levels of abstract theory only in some cases and not others. It may be, for instance, that a particular case is a simple application of a particular legal rule. Even

---

26 The literature on this topic is vast; see Dworkin (n 1 above) 36 - 48 for a discussion of some of this literature with regard to legal theory. Posner (n 13 above) 377 accuses Dworkin of failing adequately to explain what is meant by theory in law.

27 Dworkin (n 20 above) 358.

28 Dworkin (n 20 above) 358 - 359. Woolman’s charge that Dworkin somehow seeks to discover a transcendent, higher reality is puzzling given that Dworkin expressly criticises such theories in the first chapter of Justice in robes. Similarly, Woolman’s claims about how Dworkin (and those who reify theory) engage with empirical evidence would seem to be contradicted by Dworkin’s statements about the inside-out method (see below).

29 Dworkin (n 20 above) 56.
then, judges would need clear reasons for why following such a rule (or precedent) is important in a particular case and what the very value of consistency in law is. The point is that if we conceive of the relation between the concrete and the abstract in adjudication in a linear way, it is difficult to see how to avoid a continual push to ascend the theory ladder in every particular case. That would be undesirable, and possibly paralysing for a legal system that requires speedy decision making.

Secondly, in his articulation of the relation between theory and practice Dworkin constructs an opposition between two ways of proceedings. The first is that of Hercules: reasoning from the outside-in, from abstract theory to the concrete case before a judge. The second is that of ‘ordinary people, lawyers and judges’ who reason from the inside-out: ‘We begin with discrete problems forced upon us by occupation or responsibility or chance, and the scope of our inquiry is severely limited, not only by the time we have available, but by the arguments we happen actually to encounter or imagine.’ The ways of proceeding are either from the abstract to the concrete or from the concrete outwards. Yet, it seems to us a better account of adjudication would be of an inter-relationship between both methods of proceeding. Adjudication involves a backward and forward process between different levels of abstract theory and concrete facts.

In his recent book, Justice Albie Sachs gives us a glimpse at the decision-making process of a conscientious Constitutional Court judge. Sachs provocatively begins his chapter on the ‘working of a judicial mind’ with the claim that ‘every judgment I write is a lie’. By this, Sachs means that a judgment appears on its face to be logical, coherent, ‘orderly, clear, sequential’. It seems to move clearly from legal principle to concrete application in relation to the particular facts. Yet, this appearance belies the process that led to its creation which is much more chaotic:

‘the actual journey of a judgment starts with the most tentative exploratory ideas, and passes through large swathes of doubt and contestation before finally ending up as a confident exposition purportedly excluding any possibility of error.’ Sachs proceeds to contend that ‘as we wrestle with a problem, we go to-and-fro, backwards and forward, from logic to discovery, from discovery to logic’.

This account of judicial decision making suggests that the process of arriving at a decision is not linear in the way Dworkin suggests. The concrete

---

30 Dworkin (n 20 above) 54.
31 Sunstein (n 23 above) 392 admits that conceptual ascends may happen but also stresses the fact that ‘[c]onceptual or justificatory descents may well work best’.
32 A Sachs Strange alchemy of life and law (2009).
33 As above.
34 Sachs (n 32 above) 47.
35 Sachs (n 32 above) 48.
36 Sachs (n 32 above) 54.
circumstances of a case influence the way in which it is justified and the more abstract principles and theory that result. Yet, abstract theory and higher level principles also condition the way in which the concrete circumstances are viewed. This interplay between the concrete and the abstract is a productive one and a decision is arrived at when congruence exists between the different levels. The idea is less about a construction where one block is built upon another; rather, it involves the notion of a network of intersecting considerations that help justify a concrete decision.

This method of proceeding appears to be very similar to the method of justification in political philosophy that John Rawls refers to as ‘reflective equilibrium’. This method requires us first to identify what Rawls terms certain considered judgments about justice. These are judgments that are arrived at by considering ‘what one thinks about the question at hand rather than by consciously applying some principle or theory’. Such judgments are considered in the sense that they are arrived at under certain conditions: For instance, individuals will be aware of relevant facts, they are not upset or frightened, and do not stand to gain or lose by those judgments. They are also not arbitrary and are stable over time. Ultimately, these considered judgments place constraints on the theory that we will develop. For instance, if a moral theory leads to the result that slavery is just, we will rather give up the theory than the considered judgment that slavery is unjust. Considered judgments thus help to test theory. Yet, such judgments appear groundless without some form of justification.

This leads to the desire to formulate certain general principles that can account for the various judgments we make about concrete cases which cohere together. In developing those principles, we will find that they do not immediately capture all our considered judgments. The ambition would be to obtain a set of principles that can adequately account for all our convictions. In trying to achieve that ambition, we may formulate particular principles that provide a compelling account of most of our considered judgments. Such principles may, however, require us to give up some of these judgments that are not consistent with the principles. On the other hand, certain considered judgments may be so firmly held as to lead us to give up a particular set of principles that cannot adequately account for them. One is then to continue in this way, working back and forth between principles and judgments, until one reaches a set of principles and a set of judgments between which there is no conflict. This state is what Rawls calls reflective equilibrium.

37 The order of priority is also not only from concrete to abstract as Woolman suggests at 3.
38 Rawls Theory of justice (1973) 47.
40 Rawls (n 38 above) 47 - 48.
41 Rawls (n 38 above) 48.
43 Scanlon (n 39 above) 141.
Though Rawls sought to apply this method to the realm of political philosophy, in our view, it also seems peculiarly apposite to law. Indeed, we contend that it provides a better account than the notion of 'justificatory ascent' of the relation between abstract theory and the concrete practice of adjudication. Ultimately, legal theory and doctrine help develop the approach that should be adopted in concrete judgments, and concrete judgments help test whether the legal theory or principles we hold are the correct ones. This idea does not create an interminable regress to greater levels of abstract theory for it is about obtaining a 'fit' between abstract theory and concrete judgments. Moreover, it recognises the non-linear nature of legal justification which is about a backward and forward movement between the abstract and the concrete and a web of interlocking, supporting considerations.44

This account also helps explain why theory and practice in law are inevitably intertwined. And why an attempt to eschew one or the other in adjudication will be unsound.45 In the next section, we look at the notion of minimalism which attempts largely to reduce the role of theory in adjudication. Our abstract discussion thus far supports Woolman’s critique of minimalism which we bolster with a consideration of a set of concrete cases as well.

Part 2: Minimalism

Minimalism has been formulated in a rather extreme way by the South African Constitutional Court in the case of Zantsi v Council of State, Ciskei. In this case, the Court stated that judges should never ‘formulate a rule broader than is required by the precise facts to which it is to be applied’.46 This statement is, however, clearly too strong and in fact incoherent in that a rule must always necessarily be broader than the actual situation out of which it arises.47 Nevertheless, what is evident is the Court’s desire to avoid general theoretical engagement. This approach led to certain early critiques of the Court as having provided South Africa with a vague and rather superficial ‘rainbow jurisprudence’.48

Currie wrote a defence of the Constitutional Court’s method of proceeding which essentially involved defending a minimalist approach

---

44 In epistemological terminology, this is more of a coherence than a foundationalist theory. For an impressive defence of a coherence theory of justification in epistemology, see BonJour The structure of empirical knowledge (1988).
45 It should also be evident that our approach neither reifies theory nor requires a ‘theory of everything’ and thus is not subject to the critique entered into by Woolman against theorists who he claims in his introduction make these mistakes.
46 Zantsi v Council of State, Ciskei 1995 4 SA 615 para 8
47 A rule usually has general application and its application is not confined to particular persons at particular points of time. The rule of law would be impossible without such a conception of a legal rule or principle.
48 Cockrell (n 5 above).
to adjudication.\(^49\) In his article, Currie draws on the work of Cass Sunstein who has defended a minimalist practice recognising the virtues of ‘incompletely theorised agreements’.\(^50\) The insight is that people may be able to agree on particular decisions without being able to agree on a broader or abstract theory. ‘It is important because it is a way of promoting stability, reducing strains on time and capacities, and demonstrating mutual respect: it is not very respectful to take on other people’s most fundamental commitments when it is not necessary to do so.’\(^51\) Sunstein attempts to sidestep any reference to more abstract principles by stipulating an interpretive methodology based largely on the use of analogy, which he claims would not require very abstract thinking.

However, Dworkin has pointed out the error in this method of proceeding: there can be no comparison without reference to an external point. ‘Thus, for instance, and in spite of the popular proverb, one can in fact compare apples and pears provided one chooses an appropriate point of reference such as colour, taste or nutritional value.’\(^52\)

What forms an appropriate external point of reference, of course, then depends on a broader theory of the purposes of the comparison.\(^53\) Woolman points out that while Sunstein shows significant ambivalence towards theorising in judgments, he is forced to concede the inevitable importance of abstract principles in adjudication.\(^54\) In our view, Roederer has correctly argued that, at the end of day, the difference between Dworkin and Sunstein is one of attitude.\(^55\) Dworkin ultimately accepts that conceptual ascent may be limited to those instances where it is necessary.\(^56\) Similarly, Sunstein accepts the necessity of a degree of theory and has toned down his views on minimalism to become a ‘presumption, or a mood and not a rule’.\(^57\)

We have argued in the last section that there are layers of theory in the law. Judicial practice involves moving backwards and forwards between these different layers to arrive at a justifiable decision in the circumstances of a particular case. By eschewing abstract theory, a minimalist procedure removes judges from considering some of the necessary elements involved in

\(^{49}\) Currie (n 2 above).


\(^{51}\) Sunstein elaborates upon this in *Justice in robes* (n 1 above) 69.

\(^{52}\) Sunstein apparently accepts this in ‘Incompletely theorised agreements’ (n 48 above) 17 - 18.

\(^{53}\) Woolman (n 5 above) 357 - 358. Sunstein admits in ‘Incompletely theorised agreements’ (n 48) that ‘any assessment of facts is theory-laden: As Dworkin rightly stresses, we do not know what and how facts count unless we have a “normative account”’ (396).

\(^{54}\) Roederer (n 2 above) 510 - 512.

\(^{55}\) Dworkin (n 20 above) 376.

\(^{56}\) Sunstein ‘From theory to practice’ (n 23 above) 402.
legal decision making. These problems are best illustrated by focusing on some of the jurisprudence of the Constitutional Court where it has adopted a minimalist procedure: for the purposes of illustration we shall consider its case law relating to the right of access to courts.

In *Mohlomi v Minister of Defence*\(^5^8\) the Constitutional Court considered the constitutionality of a provision of the Defence Act dealing with civil actions that are instituted against the Defence Force.\(^5^9\) The impugned section stipulated that a civil action may only be launched within six months of the cause of action arising and may only be instituted once a notice of intention to commence an action is given to the defendant one month before the case is launched. The applicant had failed to serve the relevant notice a full one month before commencing the action for fear that the matter would exceed the six-month prescription limit on instituting actions. The Court had to decide whether this provision which placed time limits on when an action could be instituted infringed the applicant’s right of access to court (section 22 of the interim Constitution).

Didcott J found that what matters in deciding questions relating to the right of access to courts was ‘the sufficiency or insufficiency, the adequacy or inadequacy, of the room which the limitation leaves open in the beginning for the exercise of the right. For the consistency of the limitation with the right depends upon the availability of an initial opportunity to exercise the right that amounts in all the circumstances that characterises the class of case in question, to a real and fair one.’\(^6^0\) The judge took account of the realities of South Africa where poverty and illiteracy abound and lead to a lack of knowledge of law and access to legal representation. Consequently, he found that the right of access to courts was infringed as individuals were provided with too short a time in which to give notice and commence their claims. They had not been provided with ‘an adequate and fair opportunity to seek judicial redress for wrongs allegedly done to them’.\(^6^1\) The Court also found that there was no adequate justification for the limitation given that less restrictive means (a longer time period) could have been adopted.\(^6^2\) In analysing the right to have access to court, we find a confusion in *Mohlomi*. Rights analysis is supposed to take place in a two-stage process: The first stage involves deciding whether the right has been violated and the second stage whether the limitation on the right is justifiable. However, in *Mohlomi*, the Court, confusingly, assumes a limitation is taking place but builds a test concerning the adequacy thereof into the first stage of the enquiry. Yet, the ‘reasonableness’ and ‘justifiability’ of a limitation is precisely a matter that is supposed to be determined at the second stage of the rights enquiry. The Court appears to lack a deeper grasp of the differences between the two stages of

---

\(^5^8\) 1997 1 SA 124 (CC).
\(^5^9\) Sec 113(1) of Act 44 of 1957.
\(^6^0\) *Mohlomi* (n 58 above) para 12.
\(^6^1\) *Mohlomi* (n 58 above) para 14.
\(^6^2\) *Mohlomi* (n 58 above) para 18.
constitutional analysis and the theoretical importance of these stages. Even if it
has such a grasp, the failure to indicate the differences between them in the
cJudgment leads to confusing doctrine, a matter that foreshadows developmes
in the last case we consider.

*Mohlomi* was followed by *Beinash v Ernest and Young*. That case
considered another kind of fetter on the right of access to courts, namely,
a provision of the Vexatious Proceedings Act. This provision essentially
allowed for an application to be made to a judge to bar a particular
applicant from instituting actions in court where such an applicant had
persistently and without reasonable grounds instituted legal proceedings.
Once such an order is made, an applicant may litigate in future only if a
judge is satisfied that the proceedings are not an abuse of the process of
the court and there are *prima facie* grounds for proceeding. In this
case, the applicants had brought a significant number of cases against a
particular respondent and had been barred from bringing any further cases
unless the special circumstances envisaged in the Act were met. The
applicants challenged the constitutionality of this provision, arguing that
it unjustifiably infringed their right of access to courts (section 34 of the final
Constitution).

The Court, per Mokgoro J, dismissed the constitutional challenge to the
provision. There are a number of puzzling elements to its analysis. First, the
Court engages in constitutional analysis whilst purportedly seeking to decide
whether to grant leave to appeal. It would have seemed more sensible to grant
such leave and then engage in such constitutional analysis. This mode of
proceeding also leads the Court to conduct a more cursory analysis than may
have been required. Secondly, the Court does not engage in any detailed
analysis of the requirements of section 34: Rather, it simply moves directly
from the claim that the very purpose of the legislation was to limit the right of
vexatious litigants to access courts to the finding that a violation of section 34
had thus taken place. Notably, the judgment makes no mention of *Mohlomi*
in seeking to decide what falls within the ambit of section 34: This is a strange
omission considering that *Mohlomi* had articulated a test for when the right of
access to courts had been violated. The Court appears here to avoid
engaging with even the existing applicable law governing the area. The Court
then conducts a brief limitation’s analysis and finds that the section in question
is a justifiable limitation of section 34. Whilst the Court had placed much
emphasis on the less restrictive means component of the limitations analysis in
*Mohlomi*, it glosses over this question in *Beinash*, simply claiming that the Act
strikes ‘the appropriate balance of proportionality between means and ends’.

---

63 1999 2 SA 116 (CC).
64 Sec 2(1)(b) of the Vexatious Proceedings Act 3 of 1956.
65 Though *Mohlomi* was decided in terms of the interim Constitution, the wording
between the two Constitutions does not differ markedly.
66 *Beinash* (n 63 above) paras 17 - 21. For an earlier criticisms of this case see S Woolman
‘The right consistency’ in (1999) *SAJHR* 166.
67 *Beinash* (n 63 above) para 21.
This allows it to avoid engaging with what was perhaps the most critical issue in the case which was whether adequate procedural safeguards existed for litigants declared to be vexatious.

In prescription cases following Mohlomi and Beinash, the Court applied the ‘fair and adequate’ test laid down by Didcott J in Mohlomi and each time has struck down the provision as unconstitutional. The most recent decision dealing with a challenge to prescription periods, Road Accident Fund v Mdeyide, adds a new twist. The case concerns the constitutionality of a prescription period of three years for a claim for compensation by victims of motor vehicle accidents under the Road Accident Fund Act (RAF Act). A claim under the Act would prescribe irrespective of whether the victim knew of the identity of the negligent party and details of the facts from which the claim had arisen (unlike the provisions of the Prescription Act). The RAF Act also contains no provision for condonation of a late claim even in exceptional cases. In the facts of the particular case, a poor, disabled man was hit by a motor vehicle whilst walking along a road. For a variety of reasons, his claim was lodged three years and three days after the accident. The majority of the Court decided that whilst the prescription period violated section 34, such a violation was justifiable given its importance for the functioning of the RAF. The minority held that the violation in question was unjustifiable given the serious violation of section 34 that was involved, the social context of poverty and inequality and the possibility of less restrictive means being employed.

The reasoning of the majority in this case deserves closer attention as it is beset by several problems. First, the Court, when it deals with whether section 34 is violated, largely eschews an analysis of the right itself. In relation to Mohlomi, after outlining the approach adopted there, the Court proceeds to claim that ‘considerations taken into account during the first phase of the enquiry, could have been relevant in the second as well’. In a footnote, the Court points out that in Mohlomi ‘the fair and adequate opportunity’ test was used in the first stage of the enquiry and claims that it was used in the second stage in Brümmer. The Court claims that ‘although a two-step approach is appropriate, the questions raised and the standards applied may sometimes overlap and be applicable to both. It is not always practical to rigidly separate the two stages of the enquiry.’

68 See Moise v Greater Germiston Transitional Local Council 2001 4 SA 491 (CC); Potgieter v Lad van Uitvoerende Raad: Gesondheid, Provinciale Regering, Gauteng, 2001 11 BCLR 1175 (CC); Engelbrecht v Road Accident Fund 2007 6 SA 96 (CC); Brümmer v Minister for Social Development 2009 6 SA 323 (CC). The test was also applied in Barkhuizen v Napier 2007 5 SA 323 (CC). That case has attracted controversy for reasons that are not related to sec 34, so we will not discuss it here. See also Woolman (n 3 above) 772 - 781.

69 [2010] ZACC 18, as yet unreported.

70 Mdeyide (n 69 above) para 59.

71 Bewilderingly, this statement is blatantly incorrect. The Court cites para 51 of Brümmer in support of its statement, whereas that paragraph clearly falls within the rights analysis part of the judgment. The justification analysis, in which the test plays no significant part, follows later in the judgment from para 59 onwards.

72 Mdeyide (n 69 above) fn 52.
The Court here recognises the problem with the decision in *Mohlomi* but consigns its unsatisfactory and under-theorised answer to a footnote. In doing so, the Court calls into question the entire structure of constitutional analysis. It does so without providing an understanding as to what each respective stage involves, simply being content to blur the boundaries on the grounds of ‘practicality’. A deeper engagement with this question would have required the Court to consider the particular virtues of the two-stage process. It requires courts, for instance, to provide an interpretation of a right as well as to engage in the first stage with the violation that has taken place. Once this is clearly before the Court, a limitations enquiry can be engaged that must consider how to minimise the extent of the violation of the right even where the purpose of the limitation is justifiable. Indeed, what is notable is that by blurring the boundaries in this case, the majority leaves out any detailed analysis as to what is required at the first stage of the enquiry: It fails to engage with the interpretation of the specific right and to recognise in any detail the respects in which the impugned provision violated the right in question.

Interestingly, it is at the point where the Court is considering whether a violation of section 34 has taken place that *Beinash* suddenly enters the Court’s analysis. As was mentioned, *Beinash* provided virtually no analysis of the right of access to courts. It is thus unclear why it is referred to in deciding whether a violation of section 34 has taken place. The majority seems to use the case to justify its quick disposal of the first stage of the enquiry; however, it fails to provide any good grounds for dispensing with this stage. There is also no attempt made to reconcile *Beinash* and *Mohlomi* with the former simply standing in for the banal proposition that ‘the legislature may limit the right of access to courts in a way that is reasonable and justifiable’. The use of *Beinash* and *Mohlomi* shows the limits of analogical reasoning: Without a well-considered external reference point, the use of case law alone becomes meaningless or banal. The first stage of the enquiry is also avoided which in future may lead to a failure to give appropriate weight to the right that has been violated. What makes matters worse is that the lack of a stronger dose of theory leads the Court to fudge established constitutional doctrine concerning the two-stage enquiry and to allow resultant uncertainty in this regard to reign.

The avoidance of adequate theoretical reasoning is also evident in the approach the majority adopts towards the limitation of rights in this case. The majority recognises that they have to ‘balance the limitation of a fundamental right with the potentially calamitous consequences for the RAF, a body designed to help those who suffered as a result of road accidents, which may well otherwise follow’. In a curious footnote, Van der Westhuizen J, writing for the majority, claims that ‘[t]o weigh competing rights and governmental duties is not the same as the purely utilitarian sacrificing

---

73 *Mdeyide* (n 69 above) para 61.
74 *Mdeyide* (n 69 above) para 80.
of rights in the interest of the greater (administrative or financial) good. It is a constitutionally-mandated exercise which is well in accordance with the ideal of transformative constitutionalism'. The first sentence demonstrates the Court’s awareness of serious moral problems with a purely utilitarian process of sacrificing fundamental rights for other goods: Yet, the response is completely inept. To say that something is ‘constitutionally mandated’ does not show in any way that the balancing process differs from a utilitarian sacrifice: Such a sacrifice might be mandated by the Constitution. The simple reference to transformative constitutionalism will also not do for it is unclear in what way balancing, in itself, is transformative, particularly where rights are to be sacrificed. The judge here demonstrates an awareness of a problem for which a wholly unsatisfactory answer is provided. This does not appear to be because the judge is incapable of better reasoning but because the tropes of minimalism allow judges to make bald assertions with no adequate support.

The use of bald assertion and minimal reasoning is again in evidence in the majority’s approach to the ‘less restrictive means’ requirement. The majority considers whether the legislature could have enacted a less drastic prescription provision. Without deciding the matter, it states that the RAF and Minister believe that it could not have done so. ‘Even if they are incorrect, this is not the only consideration. The exercise is one of the proportionality test in which all the factors are weighed against one another. The mere possibility of less restrictive means is therefore not decisive.’ The Court here simply deflects the issue: By claiming that less restrictive means requirement is but one factor in the proportionality test, it fails to engage in any detail with possible alternatives to the present dispensation. Its minimisation of the importance of this element of proportionality also seriously weakens the protection that rights afford individuals in South Africa.

The majority judgment and the case law in this field demonstrate a number of the pitfalls of minimalism. First, the theory and practice of adjudication unavoidably involve a series of enquiries that exist at different levels of abstraction: The question of the adequacy of a two-stage structure is one such example that can affect the concrete outcome of a case. The minimal reasoning of the Court, in this regard, leads not only to confusion but also to significant gaps in our understanding of the way in which the rules

75 Mdeyide (n 69 above) fn 66.
76 Mdeyide (n 69 above) para 92.
77 The minority judgment in Mdeyide provides an excellent example of what more adequate adjudicative techniques would look like. Froneman J does not seek to avoid the implications of a two-stage analysis and embraces the importance of the first stage. He requires strong evidence to justify a limitation of rights and takes account of the social context in which the justifiability of a limitation must be evaluated in some detail. He also goes into more detail on the less restrictive means requirement and considers the implications of the theories of great thinkers such as John Rawls and Amartya Sen (at n 36 of his judgment) for the case at hand. For a more detailed contrast and critical engagement with the two judgments in the context of their approach to the limitation of rights, see D Bilchitz ‘How should rights be limited? Road Accident Fund v Mdeyide’ (2011) Journal of South African Law 568.
of law cohere and, thus, ultimately as to how one is to reason in relation to the Bill of Rights. The consequent adverse impact on the rule of law should be clear. As Woolman stated in his early criticism of Beinash:

My concern stems mainly from the confusion that will invariably follow from decisions that are rendered in either a haphazard or a misleading manner. Lower courts, advocates, government officials and ordinary citizens are obliged to comply with the edicts that issue from the Constitutional Court. The easier those edicts are to understand, the greater the likelihood of compliance: from High Court to lowly commoner. The Constitutional Court itself must be able to follow its own judgments. This is not a matter of stare decisis. It is, rather, a point about the building of a coherent body of jurisprudence.\(^78\)

Secondly, this lack of coherence can lead to ad hoc, ungrounded decisions. Thirdly, the Court itself avoids theoretical analysis where such reflection could counteract the applied decision it wishes to reach: Theory as a corrective upon unrestrained practical intuitions is thus rendered useless.\(^79\) Without adequate theoretical engagement, the process of reflective equilibrium cannot even get started.

We then agree with Woolman that the Court’s liaison with minimalism is undesirable for the constitutional enterprise. In the last section of this article, we turn to an assessment of Woolman’s alternative: a form of experimental constitutionalism.

Part 3: Experimental constitutionalism and minimalism

Woolman’s outline of experimental constitutionalism embraces several elements. First, it involves co-operation between different branches of government in a project of shared constitutional interpretation. Woolman claims that, over time, ‘courts, state actors and non-state actors will have the opportunity to determine whether various ‘social political experiments’ have achieved the ends set for us by our basic law (as interpreted by the courts, the legislature, the executive and non-state actors). We will, in instances of policy failure, also have an opportunity to decide whether the norms or the ends set by the courts, the legislature, the executive and non-state actors constitute the best possible gloss on our basic law’.\(^80\) Secondly, ‘participatory bubbles’ are central to experimental constitutionalism. This essentially involves setting up structures to ensure that meaningful engagement occurs concerning optimal outcomes. Broadening community participation in such institutions as the South African Human Rights Commission or even court hearings would illustrate what is meant in this context.\(^81\) Thirdly, participatory bubbles,

\(^78\) Woolman (n 66 above) 178.

\(^79\) The Court’s approach gives credence to the charge that has been levelled against judges by certain (American) Realist legal theorists and by the critical legal studies movement. For a classic exposition of the latter’s line of thinking, see eg D Kennedy ‘Form and substance in private law adjudication’ (1976) Harvard LR 1865.

\(^80\) Woolman (n 5 above) 360 (our emphasis).
Woolman claims, ‘lose their cohesion – and the pressure to produce better than zero-sum outcomes – if the courts fail to articulate the norms within which a preferred solution is meant to occur’. Fourthly, Woolman also elaborates upon the notion that experimentalism involves a continued openness to change and an ‘experimental attitude’ as to what would work best: ‘The norms created by the courts or understood by some other political body are understood to be “rolling norms” that will be subject to change where and when the exigencies of the moment require such change.’

Finally, this approach links to Woolman’s critique of minimalism when he argues that minimalism is not a virtue in itself. Woolman claims, following Sunstein, that legal rules inevitably involve shaping the social environment. This method of proceeding should involve ‘[u]nderstanding one’s environment, testing what works and what doesn’t work in that environment, putting positive results of one’s experiments into practice, being a choice architect – now that is a virtue’.

Whilst we have agreed with Woolman’s critique of minimalism and sought to develop it, we find some of his passages on experimentalism puzzling in light of the rest of his argument. First, as Woolman outlines it, experimentalism appears to involve certain procedural suggestions for developing what ‘works best’ in a constitutional system. In his enthusiasm to make Sunstein work for his argument, Woolman ultimately writes approvingly of a pragmatist method of proceeding.

Fundamental to this approach, is the idea that institutions and laws should be developed ‘from the ground up’ and according to what ‘works best’. As an account of law this is fatally flawed as it fails to recognise the way in which law is infused with a normative dimension. In the realm of engineering, if we wish to construct a bridge, it is crucial that it does not collapse. What works would thus be reasonably transparent. An experimental mode of proceeding, however, would even here be dubious given that we now have a large body of engineering theory that can help us determine in advance how to build a bridge that will not collapse. The same would appear to be true of golf. In law, however, the very notion of ‘what works’ is fundamentally under-specified and can only be filled out by normative considerations.

Dworkin, for instance, imagines a judge having to decide whether abortion is constitutionally permissible and approaching the question on the basis of ‘what works best’. For pro-choice individuals, what works best would involve enabling individuals (and particularly women) to choose

---

81 Woolman (n 5 above) 361 - 362.
82 Woolman (n 5 above) 362.
83 As above.
84 Woolman (n 5 above) 358.
85 Woolman (n 5 above) 357 - 359.
86 Dworkin (n 1 above) 63 - 65.
what to do with their bodies. For those who are pro-life, it would involve protecting the fetus. If we use another metric to measure what works such as ‘social tension’ surrounding the issue, the responses again would differ. If social tension is reduced by allowing individuals the choice to have an abortion, pro-choice supporters would see this as a sign of the success of their view. Pro-lifers would take the contrary view and see that very success in reducing social tension as a sign of the moral decay of our society. How then can what works provide a reliable guide to making decisions for judges? Ultimately, the pragmatic standard fails to provide any guidance without some prior theoretical understanding of how we measure ‘what works’. That in itself requires deep moral reflection.87

Moreover, the notion of ‘what works’ cannot be replaced by the idea of experimentalism either. For our experiments with law and ways of life need to be judged according to certain evaluative criteria. Those criteria, whilst involving an openness to change, cannot be made sense of through the notion of change or experimentation itself. Experiments in racial discrimination or forms of genocide are abhorrent. An experimental approach can only be judged as valuable if it leads to the normative world of justice that we believe to be valuable. Thus, experimentalism cannot provide a viable alternative to minimalism, for it is itself purely instrumental in nature.

The mistake Woolman makes with experimentalism is akin to the mistake he makes in his analogy between the law and golf. Theory in law is not simply a better way to improve a pre-determined outcome: It is about defining the very outcomes themselves and elaborating upon the very normative objectives of the legal enterprise. It is about interpreting texts. In doing so, we cannot be satisfied with purely instrumental methodologies that fail to grapple with the particularities of end-setting and interpretation itself.

Woolman could respond that he has explicitly stated that his focus on experimentalism is not meant to suggest that ‘all of the normative commitments in South Africa’s basic law – or the US Constitution – are up for grabs. Were the South African Constitution to set no clear normative floor, then we would not have binding constitutional law, and our Constitution would be no more than a trifle.’88 Despite this admission, Woolman does not reflect any further about what this concession means for the role of theory in law. Moreover, he does not seem to recognise that it fundamentally requires a different methodology: in the very next paragraph, he refers approvingly to Sunstein having become an empirical constitutionalist ‘when it comes to building better theories about the normative content of a country’s basic law’.89 This leaves open the question – which we have sought to emphasise –

87 As above. 88 Dworkin (n 1 above) 21. 89 Dworkin (n 1 above) 21.
as to how empiricism and an experimental approach helps in the quest to explicate normative questions.

Woolman also fails to recognise a clear tension between his approval of experimental constitutionalism and his critique of minimalism. Indeed, in laying out his version of experimental constitutionalism, he contends that 'the process of general norm-setting by the courts that initiates a process of rolling best practices by other parts of the state never gains sufficient traction when constitutional norms remain radically under-theorised'. However, the problem is that the more detailed a court’s norm-setting prescriptions are, the less room is left for experimentation and finding out 'what works best’. A defender of the minimalist approach to adjudication would argue that one of its chief virtues is precisely the use of narrow adjudicative techniques that leave open the possibility of change in the future. This incremental approach would allow one to test a particular norm in a narrow way and allow flexibility for modification in the future. The more detailed and comprehensive the norm setting of the courts the less flexibility there would be in future and the less room there would be for shared constitutional interpretation and participatory bubbles to make a difference. Experimental constitutionalism thus seems to support a minimalist approach to adjudication rather than providing an alternative to it.

Our argument is not that experimental constitutionalism as articulated by Woolman does not have a number of virtues nor that it can never provide a useful methodology in constitutional matters. However, we contend that it is an incomplete method in the realm of law and fails to provide us with a deeper understanding of the relationship between theory and practice in this domain. Recognition of the distinctive features of legal adjudicative practices is what is needed. We have argued that judges are inevitably required to engage in the realm of normative reasoning and that, consequently, the method that is required is one akin to that adopted in the realm of political philosophy where a reflective equilibrium is sought between considered convictions and theoretical understandings. New circumstances will of course disrupt an easy equilibrium and require the revision of settled doctrine. Adjudicative practices are also essentially interpretive which requires a greater degree of abstract reasoning than more applied practices. An experimental approach to institutional structures and possibly even rules may be consistent with this method. However, any account of adjudicative practice cannot be ‘experimentalist’, ‘pragmatist’, or ‘empiricist’ alone and must include an understanding of the role that normative reasoning plays in the development of law and its distinctive methodologies.

90 Dworkin (n 1 above) 25.
2 By way of conclusion: Reflection on the role of academics

We have sought to argue in this piece that the relation between legal theory and practice is not captured adequately by the analogy with the role that theory plays in golf. It is of course true that theory has a role to play in both: Yet, attending to the particularities of each practice is crucial in capturing its distinctive role in each. Of key moment in relation to law is its normative dimension as well as the fact that it involves interpretation. In light of this, we have contended that there are various layers of theory in law. Adjudication must necessarily engage some theory: We have argued that Rawls’s notion of reflective equilibrium provides a better methodology for capturing the relationship between abstract theory and concrete facts than Dworkin’s idea of justificatory ascent. As Van Hoecke argues, ‘[r]eality partly determines theory. Theory partly creates reality. This is a continuous dialectical interaction without real synthesis: It is an open-ended process.’

With this understanding of theory, what then is the role of legal academics? We do not believe that this relationship is best captured by the idea that academics in some ways ‘coach’ judges in the way Sean Foley coaches Tiger Woods. This understanding suggests a particular dynamic: The academics are those with deeper insights who are in a position to instruct judges in how to adjudicate better. The understanding is hierarchical and uni-directional. There are several problems with this account. The first is that academics are somehow solely concerned with the practice of judging and the outcomes involved. Arguably, academia itself has its own form of praxis, one that is in fact thoroughly suffused with theory. Of course, judgments play a crucial role in legal academia though academics would not necessarily see their role again as the purely instrumental one of improving judicial adjudication. There is no doubt that academics can be a form of ‘feedback mechanism’ for judges. Nevertheless, their role is more complex and would also, for instance, involve articulating general conceptions of law and its distinctive methodologies, justice and doctrine that advance the understanding of multiple actors in society. It could also involve an advocacy role for the reform of certain positions that both the legislature and judiciary are urged to take on the basis of sound justification.

91 M van Hoecke Law as communication (2002).
92 Woolman (n 5 above) 365 - 366.
93 The following claim by Woolman (n 5 above) 365 is surely overstated: ‘[W]e have only one goal: to read the Court’s judgment’s carefully, to reflect back to the Court what we see and hear, and to make the Court’s judgments better.’ This seems contradicted by Woolman’s claim (349) that ‘at its best, this knowledge system provides both the grounds for understanding the world and the conditions for offering new and better ways of being in the world’. We align ourselves more with the latter understanding of the role of academia.
94 Woolman (n 5 above) 389.
Secondly, as we have contended, law by its very nature requires an integration of theory and practice. As such, both adjudication and academic writing involve these two elements. It is true that judging has its own particularities and will be focused very much on reaching a practical result in a concrete case on the basis of sound legal justification. Academics are free to theorise without the constraints of having to reach such a result. The emphasis in adjudication is thus on the practical implications of a judgment, whilst the focus of academic writing is more squarely upon the theoretical dimensions involved. The emphasis in these two domains is thus different. The power dimension is also important: whilst the decisions of judges give rise to enforceable consequences, academic writing usually does not have such immediate consequences.

When we understand the model of legal reasoning along the lines of Rawlsian reflective equilibrium, it becomes evident that academics and judges have a mutually reinforcing role. Judges, whilst focusing on the concrete results of their decisions, place constraints on the theories that may develop and provide academics with a concrete judgment against which to test their theoretical constructions. Academics on the other hand are able to take a wider view than judges, to situate concrete cases in light of broader doctrines and theories and to point out inconsistencies and flaws in those concrete cases. We thus agree with Woolman that the role of academics may lead to a better practice of adjudication. Yet, the relationship is one that goes both ways. The backwards and forwards movement between academics and judges provides an engagement that can help avoid ungrounded theory whilst eschewing blind practice. In light of this understanding, engagements between academics and judges should be welcomed as each brings their distinctive emphasis to bear in advancing and improving the legal enterprise itself. The integration between theory and practice in law is thus reflected in the various elements of the legal profession. Both academics and judges are central to the enterprise of developing a legal system focused on achieving justice for all.
[I]nfluence-anxiety does not so much concern the forerunner but rather is anxiety achieved in and by the story ... or essay ... [T]he strong poem is the anxiety achieved ... What matters most is ... that the anxiety of influence comes out of a strong act of misreading ['poetic misprision'] ... What writers may experience as anxiety ... are the consequence of poetic misprision, rather than cause of it. The strong misreading comes first: there must be a profound act of reading that is a kind of falling in love with a literary work. That reading is likely to be idiosyncratic, and it is almost certain to be ambivalent, though the ambivalence may be veiled.

Harold Bloom *The anxiety of influence: A theory of poetry*

1 Introduction

Over the past seven years, Frank Michelman and I have engaged in an aggressive form of learning from one another: a sport New Yorkers would call ‘kibitzing’. Our first exchange around the subject matter of this colloquy (the application and the interpretation of the Bill of Rights) occurred around Thanksgiving – November 2003 – at a talk given by Professor Michelman at New York Law School.
At about the same time, I was entering my second year of writing on the Court’s ‘application’ jurisprudence and was trying to come to grips with that body of work in the post-\textit{Khumalo} era. I had taken a line, largely unwavering, since I began thinking about the subject with my late great friend Etienne Murienik. But that line – easily adopted as criticism and opposition under the interim Constitution – now had to take account of the Constitutional Court’s only pronouncement on the final Constitution’s application clause. In a unanimous judgment, the Court held that the Bill of Rights does apply directly and horizontally where the content of a right so allows, and so dictates. That conclusion was not for me my core concern. Rather, as something of a mechanic when it comes to topics such as application, interpretation and limitation, I was troubled by what I thought was Justice O’Regan’s gossamer thin five-paragraph explanation for the Constitutional Court’s take on the subject.\footnote{\textit{Khumalo v Holomisa} 2002 5 SA 401 (CC), 2002 8 BCLR 771 (CC).}

I knew what I wanted to say. The question was: How do I say it best, without being rude or intemperate? Professor Michelman, to his credit and my benefit, worked through the core of an exceedingly long rough draft of the ‘Application’ chapter. What followed was a month of intense engagement about my text, the judgment in \textit{Khumalo} and the relevant clauses of the final Constitution. During this exchange, I picked up such terms of art as ‘textual plausibility’, ‘naturalness’, ‘surplusage’, ‘range of application’, ‘prescriptive content’, and, most important of all, ‘the good faith reconstruction’. However, our exchanges did not constitute legal language lessons. They were, and still remain, part of an effort to set out the black letter law, a good faith reconstruction of the Court’s jurisprudence where the Court itself has been silent, and a preferred reading where the good faith reconstruction cannot, in my view, do the work the text naturally and plausibly requires. Throughout these exchanges, Professor Michelman continually pushed me to make my arguments leaner, meaner and more rigorous. (And he often supplied the necessary language to do so.)

What I came to learn about good faith reconstructions throughout our various exchanges was not just a technical, if important, part of my tutelage (though Professor Michelman would never describe our exchange as such). What Professor Michelman did for me is what he does for all constitutional law scholars that I know: He makes their work better. And that itself is a lesson in academic collegiality and analytical rigour – or, if, one likes, ‘aggressive learning’.\footnote{I cannot emphasise that enough. Professor Michelman is, in my opinion, one of the most influential legal academics in South Africa because of what he has done, and continues to do, for others. See, eg, H Botha \textit{et al} (eds) \textit{Rights and democracy in a transformative constitution: Festschrift for Frank Michelman} (2003).}
The chapter, and subsequent articles by Professor Michelman and myself, extended that conversation over time. This article is part of that shared effort at aggressive learning (I prefer Professor Michelman’s other apt locution – kibitzing) designed to sharpen ‘the clarity and bite of [our] differences … [while] enlarging the basis of shared opinion’.  

2 Common ground and shared opinion

Here I mark out the reworked grounds upon which we both (largely) now agree.

The first reworked ground involves a brief exposition of the place ‘the principle of charity’ occupies in Davidson’s theory of truth and interpretation. For it is a proper understanding of that theory of truth that animates our mutual effort to understand what drives the Constitutional Court and other expositors of the Court’s work.

The second reworked ground for agreement flows from my recognition that Professor Michelman’s critique of my ‘the flight from substance argument’ is largely correct: correct in the sense that we both agree that the Court has not, entirely, shirked engagement with the substantive provisions of the Bill of Rights issues in the disputes that have arisen before it. The real issue – and the heart of this article – turns on the Court’s mistaken embrace of judicial minimalism: a mistake that leads, in turn, to a lack of doctrinal clarity – in a number of areas – on the part of the Court. As Professor Michelman notes:

[Stu . . ] makes good and fair points. They go to the debate over pragmatism and minimalism. I agree that a general policy of minimalism at the Constitutional Court would be dubious for South Africa at this point. So, while I would not myself, on the whole record, characterise the Constitutional Court's performance as striking a badly wrong balance, I certainly see the strength and importance of your view.  

Both of us now hold the view, as Professor Michelman says, that ‘storming over section 39(2) and direct/indirect application – and the extent to which it leads to a flight-from-substance’ does not ‘get to the marrow’. Professor Michelman continues: ‘That was my focus in my Constitutional Court Review piece, and I can see now that it was too narrow and potentially misleading as to my overall view.’

---

4 D Davidson ‘On the very idea of a conceptual scheme’ in D Davidson Inquiries into truth and interpretation (1984) 183 197 (my emphasis).
5 E-mail from Michelman on 10 January 2010.
6 E-mail from Michelman on 12 January 2010.
7 As above.
I expect that we will still part ways on application, but that parting to me seems rather minor – even if that means that I appear incorrigible in my views on the subject. To the extent that I turn my attention to the problem of thinness, and the need for the Constitutional Court to model rational discourse, I rely on three primary sources: (1) the curial and extra-curial writings of three Constitutional Court judges – Justice Edwin Cameron, Justice (now Emeritus) Kate O’Regan, Deputy Chief Justice Dikgang Moseneke, and one High Court Judge – Judge Dennis Davis; (2) the analysis of Constitutional Court doctrine offered by fellow academics and practitioners – Cora Hoexter, Lilian Chenwi, Dario Milo, Glenn Penfold, Anton Fagan, Martin Brassey, Jason Brickhill and Frank Snyckers; and (3) my own rather piecemeal work as a legal ethnographer who worries that many judges fear constitutional analysis, or do not know how to handle a constitutional matter, because the Constitutional Court has not made patently clear the manner in which constitutional claims ought to be handled. Professor Michelman himself is a source for my argument regarding the need for greater clarity. In his critique of my position in ‘The amazing, vanishing Bill of Rights’, Professor Michelman writes:

… I have chosen to devote this space to seeing what might be said on the other side, specifically with regard to NM and Masiya, on behalf of a court that, I quite agree with Woolman, has left itself with a lot of explaining to do. The controlling opinions are indeed, as Woolman says ‘thinly reasoned’, if by that we mean they are insufficiently explained ... In developing [various] claims, I shall quite openly engage in filling in passages that may strike readers as excessively indulgent of the Court, if not entirely fanciful.8

Now, Professor Michelman’s good faith reconstructions of both NM and Masiya are not fanciful at all. What reconstructions demonstrate – convincingly enough for me to think again about some of my original contentions – is the kind of judgment that ‘ought’ to issue from the Constitutional Court if the Constitutional Court is concerned with (a) being fully understood by its broad audience of other judges, lawyers, law students and ordinary South Africans; (b) modelling, in a patient manner, rational political discourse in a contemporary South Africa sorely lacking in that domain.

The third reworked ground for agreement flows from my recognition that Professor Michelman’s critique of my application argument has more than a little merit. But I have never really denied the strength of this account. I have always admitted the obvious: that a poorly drafted text does not allow for one clear literal construction. To that extent, I have said that some interpretations of section 39(2) – such as those muscular realist readings offered by both Professor Michelman and Professor Currie – cannot simply be ignored and could, potentially, do the work that I ascribe to other sections of the constitutional text. However, to the extent that it matters, I stand firm on three propositions: (a) that clear textual markers support cabining ‘application’ issues

8 F Michelman (n 1 above) 2 - 3 citing Woolman ‘Amazing’ (n 1 above) 762 790 n 51.
in section 8 and leaving more general interpretive matters to the section that bears the name ‘interpretation’, section 39; (b) that a lack of clarity on this matter makes it difficult for non-Constitutional Court judges, practitioners and students to construct Bill of Rights arguments; and (c) that, ironically, as Sprigman and Osborne presciently pointed out over a decade ago, *Du Plessis v De Klerk* is not dead but actually lives on in the (rather confused) application jurisprudence of the Constitutional Court.

3 Davidson revisited: How best to understand the ‘principle of charity’

Recall that the real gravamen of Professor Michelman’s complaint in his article is that my penchant for ‘housekeeping’, or my ‘normative pre-commitments’ or my failure to work hard enough to understand the Court’s approach to its own application jurisprudence led me to reach conclusions not supported, in his view, by *NM* or *Masiya*. Whether the conclusions I reached were supported by the case law more generally is an issue that Professor Michelman expressly refrained from addressing.10

Professor Michelman begins his article – and his critique – with the following quote from Davidson:

Charity is forced on us; whether we like it or not, if we want to understand others, we must count them right in most matters.11

Now, as big guns go in analytic philosophy, Davidson’s principle of charity plays a critical role in one of the most powerful and best known theories about truth and interpretation. So at least, on the surface, there appears be a charge, again limited to two texts (*NM* or *Masiya*), that I have violated a widely accepted canon of engagement with the language and the texts of others. Perhaps the charge is far weaker – and does not amount to a charge that I have made one or more philosophical errors. Perhaps the charge flows from (a) the degree of censoriousness with which [Professor Woolman] has approached the Constitutional Court’s work [in *NM* or *Masiya*]; and (b) the weight [Professor Woolman] place[s] on niceties of the “application” debate.12

---


10 Michelman (n 1 above) 2.
11 Davidson (n 4 above) 197.
12 E-mail from Michelman on 29 December 2009.
I want to pry apart, in the next few pages, the use of the ‘principle of charity’ – which I think has been misunderstood and may continue to be misunderstood – from the critique that my previous writing had been overly censorious and contingent upon an application argument that does not do all the work that I claimed it does. This claim I now accept. My argument runs as follows: The principle of charity is (a) not so much a tool (for ethnography) (b) so much as it is an important feature of a much larger argument (c) about how we understand the construction of language, truth conditions, and the validity of philosophical and anthropological arguments that cast significant doubt about relativism of many kinds.

Before I turn to the finer points of that argument, it’s important to note what Professor Michelman did and did not claim for his Wikipedian version of the principle of charity. First, we are, he says, engaged in a 'lawyers' kibitz' over a range of Constitutional Court doctrines. Second, Professor Michelman seems to me to be on firmer ground when he returns from Wikipedia and identifies Davidson as ‘one of the principle’s chief philosophical architects and expositors’, that Davidson also calls it ‘the principle of rational accommodation’, and that Davidson summarises his position as follows: ‘We make maximum sense of the words and thoughts of others when we interpret in a way that optimises agreement.’

I disagree with the last statement – ‘that the philosophical underpinnings … need not delay us’. I have already heard ‘the principle of charity’ used, incorrectly, in legal discourse in South Africa. It’s enticing to be able to say that you are working with the ‘principle of charity’ and to be under the misapprehension that it has something to do with charity per se when it comes to a legal argument.

13 Michelman (n 1 above) 153 citing D Davidson ‘Thought and talk’ in Davidson (n 4 above) 197.
14 Michelman continues: ‘All that matters from that department is the motivation for adherence to the principle, which the moniker ‘charity’ does not very well convey. The aim of interpretive charity is not generosity toward others, or anything like that. It is not to pay homage, deference, or respect to our interlocutors, or to avoid giving offense. It is not to demonstrate our own good manners, or to toe some Goody Two-Shoes line against critiques that are not ‘constructive’. (I hold Stu Woolman’s pull-no-punches style of court-watching to be entirely constructive and admirable.) No, the aim of ‘charitable’ interpretation is not any of those. The aim is to learn. It is aggressively to learn what there is to be learnt from puzzles the interlocutors pose to us, by assuming there is method in their madness and doing our best to ferret that out, using everything else we know or can guess (in part from their likeness and kinship to us) about where they are coming from. ’To see too much unreason on the part of others’, Davidson says, is ‘to undermine our ability to understand what it is they are so unreasonable about.’ It is to risk missing issues that might merit our consideration.’ Michelman (n 1 above) 4 (citations to Davidson omitted, my emphasis).
What drives our collective preoccupation with the principle of charity? In short, the principle reflects our ability to translate the words of even the most obdurate interlocutors into propositions that we can understand, and, furthermore, confidently credit as being true. Davidson strikingly suggests that it makes no sense to speak of a language. If we can translate the words and the actions of fellow speakers well enough to support that claim, then it makes no sense to distinguish between different linguistic schemes. If that assertion withstands inspection, then the epistemological claims of conceptual relativism do not survive careful scrutiny. Thus, we can see an ethical imperative at work: We owe our fellow human beings an obligation, as subjects of equal concern and respect, to make the best possible sense of what they are trying to convey to us.

3.1 ‘Thought and talk’

Donald Davidson is concerned with the truth. And the principle of charity can be understood only once ‘truth’ is placed at the centre of things. In ‘Thought and talk’, Davidson writes:

One thing that gradually dawned on me was that while Tarksi intended to analyse the concept of truth by appealing ... to the concept of meaning (in the guise of sameness of meaning or translation), I have the reverse in mind. I considered truth to be the central primitive concept ... Something else that was slow coming to me was that since I was treating theories of truth as empirical theories, the axioms and theorems had to be viewed as laws. So a theorem like ‘Schnee ist weiss’ is true in the mouth of a German speaker if and only if ‘snow is white’ has to be taken not merely as true, but as capable of supporting counterfactual claims. Indeed, given that the evidence for this law, if it is one, depends ultimately on certain causal relations between speakers and the world, one can say that it is no accident that ‘Schnee is weiss’ is true if and only if snow is white; it is the whiteness of snow that makes ‘Schnee is weiss’ true.15

Snow is ‘white’ – because it is white in the world – in English or Sepedi or German or Zulu – and not because it is ‘white’ relative to a particular linguistic set of conventions. Of course, that does not mean that our statements about the world cannot be ‘false’. As Davidson powerfully puts the point:

But of course it cannot be assumed that speakers never have false beliefs. Error is what gives belief its point. We can, however, take it is as given that most beliefs are correct. The reason for this is that a belief is identified by its location in a pattern of beliefs: it is this pattern that determines the subject matter of the belief, what the belief is about. Before some object in, or aspect of, the world can become part of the subject matter of a belief (true or false) there must be endless true beliefs about the subject matter.16

---

15 D Davidson ‘Introduction’ in Davidson (n 4 above) xiv (my emphasis).
16 Davidson (n 13 above) 155, 168 (my emphasis).
‘Endless true beliefs’. If I had to pick out one phrase in the philosophical canon that regularly does some theoretical heavy lifting, it would be ‘endless true beliefs’. It is this notion of ‘endless true beliefs’ that allows us – today – to engage texts several millennia old, or a variety of twenty-first century South African texts, across Zulu, Sepedi, Afrikaans, Xhosa and English, or bodies of constitutional law in a variety of polities, say Germany, India and Canada. Put somewhat differently, Davidson writes:

> It isn’t that any one belief necessarily destroys our ability to identify further beliefs, but the intelligibility of such identifications must depend on a background of largely unmentioned and unquestioned true beliefs ... What makes interpretation possible then is the fact that we can dismiss a priori the chance of massive error. A theory of interpretation cannot be correct that makes a man assent to very many false propositions: it must generally be the case that a sentence is true when a speaker holds it to be true. But of course the speaker may be wrong; and so may the interpreter.17

The principle of charity forces us to recognise that ‘most’ of our fellow human beings are generally correct about ‘most’ of the statements that they make about the world. Davidson notes that

(a) ‘this working assumption actually allows us to sharpen areas of disagreement – so that we might later arrive at more precise propositions about truth’;18
(b) ‘minimising disagreement or maximising agreement is a confused ideal. The aim of interpretation is not agreement but understanding. My point has always been that understanding can be secured only by interpreting in a way that makes for the right sort of agreement. The “right sort”, however, is no easier to specify than to say what constitutes a good reason for holding a particular belief’.19

I think that Professor Michelman has, unwittingly, led some readers to believe, wrongly, that we (two authors and our audience) are engaged in the kind of inter-lingual interpretation (across Gadamerian-like horizons of space or time) that requires the principle of charity, when, in fact, we all speak English and are all familiar with the text of the basic law, its history, its setting, and its construction by the Constitutional Court and all active commentators. The principle of charity is only necessary when we start from a position of ‘radical’ difference in language that blocks understanding. That kind of difference in language or understanding does not obtain in this extended kibitz about the meaning of Constitutional Court’s judgments. Indeed, the proximity of our positions – even when it comes to occasional differences in style – is what permits us to close the circle of agreement through extended conversation.

17 Davidson (n 16 above) 155, 168 - 169 (my emphasis). Here is the takeaway: In isolated cases a speaker might be mistaken about the state of objective reality, but it does not follow that the entire body of knowledge possessed by the speaker is undermined.
18 Davidson (n 15 above) xvii.
19 As above.
3.2 ‘On the very idea of a conceptual scheme’

Philosophers of many persuasions are prone to talk of conceptual schemes. Conceptual schemes, we are told, are ways of organising experience: they are systems of categories that give form to the data of sensation; they are points of view from which individuals, cultures or periods survey the passing scene. There may be no translating from one scheme to another, in which case the beliefs, desires, hopes, and bits of knowledge that characterise one person have no true counterparts with respect to the subscriber to another scheme. Reality is relative to a scheme: what counts as real in one system may not in another.20

Davidson goes on to write: ‘Conceptual relativism is a heady and exotic doctrine, or would be if we could make sense of it. The trouble is, as so often in philosophy, it is hard to improve intelligibility while retaining the excitement.’21 One of Professor Michelman’s truly extraordinary (Davidsonian) gifts is to be able to show how writers making apparently disparate claims are actually singing off the same hymn sheet. Ought he and I to be understood to be singing off the same hymn sheet? Or is one of us tone deaf to what the Court has actually said in *NM* and *Masiya* in a manner that might have closed down the area of disagreement or at least demonstrated, more convincingly, where the real problem lies? (Naturally, upon reflection, I incline towards the former position – while leaving space for minor, but still meaningful disagreement.)

So let us return again to Davidson, and understand why Professor Michelman turned to him in the first place. Davidson writes about the principle of charity as follows:

Since knowledge of beliefs comes only with the ability to interpret words, the only possibility at the start is to assume general agreement on beliefs. We get a first approximation to a finished theory by assigning to sentences of a speaker conditions of truth that actually obtain (in our opinion) just when the speaker holds those sentences true. The guiding policy is to do this as far as possible ... The method is not designed to eliminate disagreement, nor can it; its purpose is to make meaningful disagreement possible, and this depends entirely on a foundation – some foundation – in agreement ... Since charity is not an option, but a condition of having a workable theory, it is meaningless to suggest that we might fall into error by endorsing it. Until we have successfully established a systematic correlation of sentences held true with sentences held true, there are no mistakes to make. Charity is forced upon us; whether we like it or not, if we want to understand others, we must count them right in most matters. If we can produce a theory that reconciles charity and the formal conditions for a theory, we have done all that can be done ... We make maximum sense of the words and thoughts of others when we interpret in a way that optimises agreement ... Where does that leave the case for conceptual

---

20 Davidson (n 4 above) 183.
21 As above.
relativism? The answer is, I think, that we must say much the same thing about differences in conceptual schemes as we say about differences in belief: we improve the clarity and bite of declarations of difference, whether of scheme or opinion, enlarging the basis of shared (translatable) language or of shared opinion ... In giving up dependence on the concept of an uninterpreted reality, something outside all schemes and science, we do not relinquish the notion of objective truth – quite the contrary. Given the dogma of a dualism of scheme and relativity, we get conceptual relativity, and truth relative to a scheme. Without the dogma, this kind of relativity goes by the board ... In giving up the dualism of scheme and world, we do not give up the world, but re-establish unmediated touch with the familiar objects whose antics make our sentences and opinions true or false.22

Proposition 1: ‘Conceptual schemists’ tend to focus primarily on the margins of thought – where difference is, naturally, more likely to occur. Davidson turns our attention back to the simple truth that we all believe most of the same things about the world: ‘that is the sun’; ‘that is a tree’; ‘snow is white’; ‘human beings are born, live and die’. The web of simple truth propositions about the world that exists in various populations – linguistic, ethnic, national, and dare I say even religious – is universally shared.

Proposition 2: ‘The method is not designed to eliminate disagreement, nor can it; its purpose is to make meaningful disagreement possible, and this depends entirely on a foundation – some foundation – in agreement.’23 We do not need full agreement – nor can we get complete agreement on various propositions. Our knowledge of the world is vast but finite – and at the margins where we are still trying to solve various problems, disagreement will occur.

Proposition 3: ‘Without the dogma, … relativity goes by the board. … In giving up the dualism of scheme and world, we do not give up the world, but re-establish unmediated touch with the familiar objects whose antics make our sentences and opinions true or false.’24 Or as Mamet puts it: ‘There. Is. A. Real. World. Out. There.’ About that Professor Michelman and I are certainly in agreement: Only that understanding of truth and the world has allowed our kibitzing to continue over time.

Why then so much ink spilled on Davidson by Professor Michelman and myself? Because charity, properly understood, is not the issue here. Clarity is.

22 Davidson (n 4 above) 196 - 198.
23 Davidson (n 4 above) 196 - 197.
24 Davidson (n 4 above) 198.
4 How best to understand the case for clarity

One can describe the problem of thinness in South African constitutional law in a number of different ways. Here is how I would state the problem: Thinly reasoned judgments presuppose that the current information deficit in deciding a matter – and the concomitant limits placed upon doctrinal development – will at least partially be ameliorated as time and experience throw up new opportunities to expand our understanding of how specific substantive rights ought to function in given environments. The Constitutional Court’s ongoing failure to develop coherent doctrines in a significant number of areas of fundamental rights jurisprudence places the Court’s very authority at risk. Moreover, the absence of clearly articulated rules undermines rational political discourse. Reasoned disagreement can only take place when parties agree on the general terms of the debate. The Constitutional Court must, in terms of its institutional role, establish the meaning of constitutional norms and thus the general framework for political contestation. The Constitutional Court abdicates this institutional responsibility to model rational political discourse by refusing to state, in a comprehensive manner, the reasons that ground its conclusions.

Professor Michelman is unlikely to concur with this assessment of the Court’s work. In any event, neither of us has claimed nor felt it necessary to defend the Court against the claim that it has systematically failed to offer reasoned accounts for its decisions. The former position would be ludicrous; the latter position unnecessary. Instead, in the pages that follow, I want to allow other voices to break in to this conversation – from the bench, from the podium, from the academy – to offer their assessment on the thinness and the clarity of the Court’s judgments.

4.1 Voices from the bench

Constitutional Court Justice Edwin Cameron has agreed, at least in part, with my rather abstract characterisation of some of the Constitutional Court’s judgments as thin, and lacking in clarity.25 Cameron J (then of the Supreme Court of Appeal in 2009) writes:

At this tender stage of our legal development, the doctrine of precedent has special importance. The CC has been accused of disregarding its own decisions without convincing reason (without, indeed, acknowledging that it has done so). That is a grave charge … [And] this means that other courts, including the SCA, must follow

the binding basis of its decisions in all cases in which it has assumed jurisdiction.\textsuperscript{26}

He also cites a broad array of contradictory administrative law judgments in the lower courts as evidence of the deleterious consequences of Constitutional Court judgments so thinly reasoned that they lead to contradictory outcomes in the lower courts.\textsuperscript{27} Cameron J’s critique – occurring as it does in his last judgment on the Supreme Court of Appeal – is subtly cast: The Supreme Court of Appeal cannot be seen to remonstrate (too often) the Constitutional Court on constitutional matters. Institutional comity demands a muted response.

Justice O'Regan’s minority judgment in \textit{CUSA v Tao Ying Metal Industries (Pty) Ltd; Commission for Conciliation, Mediation and Arbitration and The Metal and Engineering Industries Bargaining Council} provides further grist for the mill.\textsuperscript{28} The majority justified its ‘use’ of constitutional values to dispose of the matter in a single sentence at the very end of the judgment. Ngcobo J wrote:

In my view, the meaning preferred in this judgment accords better with \textit{the values of our Constitution}. This is so because, on the facts of this case, a labour regime that enabled the greater exploitation of black people in the homelands as part of the apartheid scheme, to the detriment of the workers in these areas, would, on the employer's interpretation, be kept in force for longer than the interpretation preferred in this judgment.\textsuperscript{29}

O'Regan J has two primary complaints about the majority’s rather loose use of ‘values’ to determine outcomes, especially when it is not even clear that a constitutional matter has been raised. Justice O’Regan first notes:

In this case, the primary dispute insofar as it relates to the bargaining council agreement turns on whether the wage provisions of the 1998 Main Agreement apply to the employer or whether the exemption granted on 7 April 1997 exempts the employer from those provisions. This does not seem to me to raise a constitutional matter. There is no provision in the Constitution which is directly relevant to the interpretation of either the 1998 Main Agreement or the exemption; nor can it be said that either of the interpretations for which the parties contend gives greater or lesser effect to the provisions of the Bill of Rights. I should add that the exemption itself at which the interpretive debate is really directed is not ‘legislation’ that falls to be interpreted in a manner consistent with the spirit, purport and objects of the Bill of Rights. Ngcobo J suggests that the enforcement of collective

\textsuperscript{26} In \textit{True Mhadi} (n 25 above) para 102 fns 52 & 53, Cameron J cites, in support of his claim, ‘Constitution chapter 1, founding provisions, sec 1 – the Republic of South Africa is founded on values that include… (c) Supremacy of the Constitution and the rule of law’… and Woolman “The amazing, vanishing Bill of Rights” (2007) 124 \textit{SALJ} 762 -794'.

\textsuperscript{27} \textit{True Mhadi} (n 25 above) para 102, n 53.

\textsuperscript{28} 2009 2 SA 204 (CC), 2009 1 BCLR 1.

\textsuperscript{29} \textit{CUSA} (n 28 above) para 103 (my emphasis).
agreements is crucial to a society founded on the rule of law. I agree. I do not think, however, that the consequence of this assertion is that the enforcement of all collective agreements automatically raises a constitutional matter, for the reasons I have given above. The rule of law of course requires that all binding obligations be enforced. It does not mean, however, that the enforcement of all binding obligations necessarily raises a constitutional matter. The 1995 Labour Relations Act carefully provides procedures to ensure that collective agreements are enforced. Those procedures have not been challenged as inadequate or unconstitutional. This Court should recognise that the Constitution establishes it as a court that has jurisdiction in constitutional matters only; not as a general court of appeal in all matters. This Court must respectfully observe those limits placed on its jurisdiction.30

However, if the Constitutional Court is going to turn an issue into a constitutional matter, then, O’Regan J contends, it has an obligation to do more than hand-wave toward its powers under section 39(2) of the Constitution when rendering a decision:

Finally, Ngcobo J states in one sentence that the exemption should be read in the way he proposes ‘because, on the facts of this case, a labour regime that enabled the greater exploitation of black people in the homelands as part of the apartheid scheme ... would ... be kept in force for longer’. With respect, this argument ignores the undisputed facts on the record before us. Those facts make plain that when the leather industry required leather companies in Botshabelo to pay the wages provided for in its main agreement and did not issue exemptions, all the leather companies in Botshabelo closed, with the devastating consequence of job losses. Accordingly, when the Department of Labour realised that the metal and engineering industries bargaining council was about to register employers in Botshabelo, it requested the bargaining council to explore ways of ensuring that businesses were not forced to close as a result of being required to pay wages beyond their means. It was at least partly as a result of this intervention that the council appointed a team of investigators to visit the employer’s premises and, in particular, to investigate the financial circumstances of the employer. It was after this visit that the exemptions were granted. In the light of the intervention by the Department of Labour, it seems likely that the exemptions were granted to avoid further job losses in Botshabelo. In my view, the one-sentence argument relied upon by Ngcobo J ignores this complex economic and social background. This Court should be slow to base its reasoning on such arguments, particularly when they have not been raised either by the union or the employer, and when they are likely to mask complex social and economic realities, possibly with harmful consequences, such as job losses.31

Read together, O’Regan J’s remarks unmistakably identify the dangers of ‘thinly reasoned’ reliance on ‘constitutional values’ to dispose of ‘complex social and economic realities’. Here, the issue for Justice O’Regan is not one of ‘charity’, but one of ‘clarity’.

30 CUSA (n 28 above) paras 127 - 128.
31 CUSA (n 28 above) para 148.
Judge Dennis Davis has laid a similar complaint. He has stated, in a public forum, that the absence of rule-based or substantive content in various areas of the Constitutional Court’s Bill of Rights jurisprudence quite often makes it difficult for High Courts to discharge effectively their function.32

4.2 Voices from the academy

While I am certain that opposing voices to my position exist in the academy, it is worth noting that a decided majority of the academics with whom I engage, and whom I read, agree that the problem facing the Court and its interpreters is not one of charity, but clarity. In Clearing the intersection? Administrative law and labour law in the Constitutional Court, Hoexter introduces the Court’s treatment of this intersection as follows:

As I shall show in this article, the results are … rather disappointing, at least from the perspective of an administrative lawyer. In my commentary on these three cases (Sidumo, Chirwa and Masetlha) that is the perspective I adopt. My aim is to … make three main points. First, it seems to me that from the various judgments of the Constitutional Court no coherent picture emerges of this intersection or of the relationship between the rights concerned. Indeed, in their constitutional approach or ethos the Sidumo and Chirwa majority judgments seem quite different, notwithstanding that the two cases were decided by almost identical benches and within weeks of each other. A theme emerging strongly from Sidumo, and to a lesser extent from Masetlha, is that fundamental rights cannot be hermetically sealed from one other. In Sidumo labour law and administrative law intermingle and converge, and in Masetlha administrative law in the broad sense is permitted to solve a problem that labour law apparently fails to address. In Chirwa, by contrast, the majority insists on the strict compartmentalisation of fundamental rights and on the pre-eminence of one right to the exclusion of the other. Sections 23 and 33 of the Bill of Rights are placed in separate jurisdictional boxes, and the attitude seems to be that labour law and administrative law must have nothing to do with one other. Secondly, in Chirwa the various approaches to the ‘administrative action’ issue are all problematic to some extent, and the majority view on the question of jurisdiction is wholly unconvincing. While the Court certainly managed to tidy up the busy intersection of labour law and administrative law in this case, it did so at the expense of the Constitutional Court’s own precedent — and apparently without much faith in the success of its efforts. Thirdly, whatever labour lawyers may think of the majority judgment in Masetlha, administrative lawyers are likely to be perturbed by it. For us it is a decision that appears to set the law of procedural fairness back twenty years.33

Hoexter finishes with this flourish:

---

32 See D Davis ‘Democracy, dignity and deliberation’ Conference on dignity and the jurisprudence of Justice Laurie Ackermann (University of Cape Town 27 July 2007).
[T]he judgments of the Constitutional Court in these cases also leave one with a sense of wasted opportunity. *Maselha* seems a retrogressive decision, at least to an administrative lawyer, for allowing an unfortunate exception to the established principles of procedural fairness. And *Sidumo* and *Chirwa* appear to be far apart in the constitutional vision or ethos expressed by the majority in each case, for the former encourages the interconnectedness of rights while the latter disavows it. As I have argued above, *Chirwa* is unsatisfying in other respects as well: none of the Court's approaches to the 'administrative action' issue is unproblematic, and the majority's reasoning on the question of jurisdiction is not only unconvincing but also contradicts the Court's own jurisprudence. The decision in this case certainly clears up the intersection between labour law and administrative law, but the cost of that tidiness may be thought unacceptably high.34

These cases – when read with *Fredericks* – create a miasma of constitutional labour law and constitutional administrative law: just as Professor Hoexter predicted. Moreover, the above account cannot be dismissed as one lone academic's opinion. In the recent decision of *Gcaba v Minister of Safety and Security*, the Constitutional Court acknowledged that a significant number of High Court and Supreme Court of Appeal cases (the Constitutional Court lists over 30) had gone in opposite – and conceptually inconsistent – directions because the jurisprudence of the Constitutional Court had been both radically under-theorised and contradictory in conclusion.35 Van der Westhuizen J opens *Gcaba* with something of an apology for sowing the seeds of this confusion:

One of the purposes of law is to regulate and guide relations in a society. ... Yet the legislature, courts, legal representatives and academics often create complexity and confusion rather than clarity and guidance. In the case of fairly new legislation based on a young Constitution this is perhaps understandable. Sometimes a jurisprudence needs to develop along with the insight and wisdom emerging from a debate over some time. The legislature may also have to intervene in appropriate circumstances, for example, when incremental development results in uncertainty or an otherwise unsatisfactory situation.36

One must give the *Gcaba* Court some credit for identifying a problem and taking responsibility for it. The Court then accepts the need to correct its course: 'One of the purposes of law is to regulate and guide relations in a society ... Yet the ... courts ... often create complexity and confusion rather than clarity and guidance.'37

---

34 Hoexter (n 33 above) 234.  
36 *Gcaba* (n 35 above) paras 1 & 2.  
37 *Gcaba* (n 35 above) para 1 (my emphasis).
The question, of course, is whether they have succeeded in changing course.38

Penfold and Milo offer their own meditation on a lack of clarity with respect to the Court’s ostensible development of the element of dolus (intention) in an actio injuriarum – and they do so with respect to one of the two cases with which Professor Michelman himself engages: NM. Indeed, they not only engage NM – they engage Professor Michelman’s essay itself:

Elsewhere in this volume, Michelman, adopting what he describes as a ‘charitable’ reading of the majority judgment in NM v Smith, suggests that the majority may have developed the common law ‘in response to constitutional pressure’ and that all of the issues in this case stood to be decided in the shadow of the Bill of Rights and of the looming possibility that the common law might, depending on how they were to be decided, have to be developed under the mandate of section 39(2). Michelman goes on to assert that the majority, in finding that the conduct of the defendants amounted to intention, may be said to have been considering whether the common law rule applies to the new factual situation as envisaged in K v Minister of Safety and Security. While Michelman’s thoughtful argument merits a more detailed response than this case note permits, it seems to us that there are at least three possible responses. The first is that one cannot get away from the fact that a finding of intention (dolus) involves a finding of subjective state of mind (ie actual knowledge or foresight) on the part of the defendant. This is a purely factual finding which, it seems to us, leaves no space for a consideration of constitutional rights or values. Second, if the majority intended, in light of the Bill of Rights, to extend the scope of dolus eventualis to conduct which would otherwise not have constituted this form of intention, one would expect them to have said so expressly. On the contrary, the approach of the majority was simply to apply the dolus eventualis test to the facts of the case before them. Third, even if one is able to bring the majority’s approach within the wording of O’Regan J’s judgment in K v Minister of Safety and Security, it seems to us more difficult to do so in relation to Luiters. Irrespective of whether or not the majority was correct in holding that the defendants’ intention (or lack thereof) properly fell within the Court’s jurisdiction, our primary difficulty with the judgment of the majority in NM v Smith is that none of these issues were canvassed in the judgment and no attempt was made to reconcile the majority’s approach with the previous case law. The Court thus dealt with a matter that could conceivably fall beyond its jurisdiction – an outcome that would have serious implications for the rule of law – without providing a meaningful explanation of its reasons for doing so.39


Although Penfold and Milo and Michelman clearly disagree on a number of points – whether the Court rightfully exercised jurisdiction or legitimately developed dolus in an appropriate manner – that is not of interest to me.\textsuperscript{40} Even if Professor Michelman claims that the Court could exercise jurisdiction or develop the scope of dolus eventualis, and Penfold and Milo demur, what is interesting is the point of agreement between the authors. And that is this: Both papers criticise the majority in NM for the thinness of the judgment. It is that thinness that drives Professor Michelman to ‘fill-in’ the judgment in the manner that he does.

Chenwi is not nearly so sanguine about the prospects of ‘filling in’ the Court’s recent socio-economic rights decisions: For Chenwi, \textit{Occupiers of 51 Olivia Road v City of Johannesburg} operates as an archetypal example.\textsuperscript{41} In \textit{Occupiers of 51 Olivia Road}, the Court was asked to determine whether the City of Johannesburg had the power to evict persons from unsafe dwellings without first providing alternative accommodation. The Court never actually established a norm that might direct other municipalities faced with the same dilemma toward a particular outcome. Instead, it directed the parties to redirect their energies toward a settlement with which all interested stakeholders could live. That they did. And the Court, in its judgment, ratified that settlement. Again, by simply ratifying the settlement, the Court provided no framework within which all parties to future, similar housing disputes might work. Chenwi contends that the thinness of the judgment in \textit{Occupiers of 51 Olivia Road} neither promotes a dialogue between the three branches of government about the meaning of section 26 of the Constitution, nor truly entertains the proposition that citizens might have a meaningful voice in the construction of our basic law.\textsuperscript{42}

\textsuperscript{40} But that long paragraph is not the only place in which Penfold and Milo reject Michelman’s reading of NM. Elsewhere Penfold and Milo write: As Woolman [in Woolman ‘The amazing, vanishing Bill of Rights’ (2007) 124 SALJ 762, 783] comments: ‘The majority in NM acts as a trier of fact in a run-of-the-mill actio iniuriarum matter.’ It may be suggested that the approach of the majority in NM falls within the second instance in \textit{K v Minister of Safety and Security}, ie determining whether a particular set of facts falls within or beyond the ambit of an existing rule, in that the majority determined that the facts fell within the ambit of dolus eventualis in circumstances which would, in the absence of the Constitution, amount only to negligence. It may be suggested that the approach of the majority in NM falls within the second instance in \textit{K v Minister of Safety and Security}, ie determining whether a particular set of facts falls within or beyond the ambit of an existing rule, in that the majority determined that the facts fell within the ambit of dolus eventualis in circumstances which would, in the absence of the Constitution, amount only to negligence. For a variation on this argument, see ... Michelman’s ‘On the uses of interpretive “charity”,’ ... For the reasons set out in this article, we do not find this argument persuasive.’ Penfold & Milo (n 39 above) 319.

\textsuperscript{41} 2008 5 BCLR 475 (CC).

\textsuperscript{42} L Chenwi ‘A new approach to remedies in socio-economic rights adjudication: 51 Olivia Road v City of Johannesburg’ (2009) 2 Constitutional Court Review 371.
For Snyckers, the thinness of the Court’s criminal procedure jurisprudence – as evident in *Molimi, Zuma, Thint (Holdings), Shaik and Zealand* – imposes a different set of costs on a single set of parties: the accused.43 Although the Constitutional Court certainly got off to a fast start in protecting the accused with respect to its reverse onus decisions, Snyckers contends that recent cases suggest a weakening of that resolve. He traces that weakening to the Court’s preference for determining specific cases in terms of the balancing of rights and interests rather than in rule generation in terms of the panoply of criminal procedure protections found in the various substantive provisions of the Bill of Rights. He writes:

In an adversarial criminal justice system, rules tend to favour the accused. Rules, when applied strictly, create rights. The fewer the rules, the more likely the conviction. The longer the view taken by a justice system, the more generous the system to those at its barrel-end. Due process intrudes upon the immediate desire to punish with an appeal to the perennial need to be humane. It arrives at the critical moment, embodying the conscience of society, to spoil the quenching of the bloodlust at the hanging party ... More often than not, [the accused’s] salvation will lie in the extent to which he is able to invoke rules, rights and principles that were created for the benefit of others, for situations other than his, and for the long-term benefit of society.44

Naturally, Snyckers continues, the accused will prefer rather uncritical application of these rules rather than nuanced assessments of the application of the rule in a particular set of circumstances. But the thwarted desires of the accused are not what troubles Snyckers here. His concern – highlighted by the quintet of cases handed down by the Constitutional Court in 2008 (*Molimi, Zuma, Thint (Holdings), Shaik and Zealand*)45 – is that the Court’s reflexive invocation of society’s interests in fighting crime has a tendency to outweigh the rights-based or rules-based arguments of the defendant. Or as Snyckers puts it: ‘Such balancing exercises will tend, in the nature of things, to end badly for [the defendant].’46 And that is so because the Court has a tendency to engage in hand-waving exercises when it comes to the content of a right – too often preferring a notional approach to rights analysis (in which the abrogation of the right is merely assumed) as opposed to a value based rule-generating

43 F Snyckers ‘The flight from rights: Rule aversion in dealing with the criminal process *Molimi, Zuma, Thint (Holdings), Shaik and Zealand*’ (2009) 2 Constitutional Court Review 269.
44 Snyckers (n 82 above) 1.
45 *S v Molimi* 2008 3 608 (CC); *Thint v National Director of Public Prosecutions, Zuma v National Director of Public Prosecutions* 2009 1 SA 1; *Thint Holdings v National Director of Public Prosecutions, Zuma v National Director of Public Prosecutions* 2009 1 SA 141; *S v Shaik* 2008 5 SA 354 (CC); *Zealand v Minister of Justice and Constitutional Development* 2008 4 SA 458 (CC).
46 Snyckers (n 43 above) 1.
approach to rights in which the content of the right must be fleshed out.\textsuperscript{47} In such cases, the Court is not long detained by the right invoked and proceeds rather immediately to limitations analysis. Does Snyckers’s argument service the case against thinness? A Constitutional Court averse to rule-generating decisions and more inclined to decisions that stick to the facts of the case is invariably going to produce ‘thin’ judgments. And by thin, again, I simply mean short on the kinds of reasons that can be extended and applied in future cases.

The final two academic witnesses in the case against thinness are Professor Michelman and myself.

Can the cases we address – \textit{NM, Masiya} and \textit{Barkhuisen} (amongst others)\textsuperscript{48} – service a more modest argument regarding thinness? They can. The argument runs like this:

\begin{quote}
[T]he Court often … substitutes … dignity, equality and freedom [value analysis] for … the more rigorous interrogation of constitutional challenges [that should be] required in terms of the specific substantive rights found in chapter 2 of the Constitution. If the drafters of the Constitution had intended such a substitution, the structure and the language of the Bill of Rights would have reflected that intention. It doesn’t … By continually relying on section 39(2) of the Constitution to decide challenges both to rules of common law and to provisions of statutes, the court obviates the need to give the specific substantive rights in chapter 2 the content necessary to determine the actual validity of the rule being challenged in the instant matter and of similar rules challenged in subsequent matters. This strategy also enables the court to skirt the nuanced process of justification that section 36 … or some other express limitations clause in a specific substantive right might require. The persistent refusal to give rights identifiable content, by avoiding direct application, results in a Bill of Rights [with less identifiable content and] meaning than it should contain.\textsuperscript{49}
\end{quote}

\textsuperscript{47} Snyckers (n 82 above) 1 - 2, n 2. (This article echoes ‘(at times loud[ly]) the lament issued by Woolman in Woolman ‘The amazing, vanishing Bill of Rights’ …). The echoes are at their loudest in Woolman’s complaint that the Court’s liking for outcomes based jurisprudence could, at its logical extreme, amount to a violation of the rule of law, or what is still sometimes called palm tree justice, at the cost of institutional legitimacy. Two phenomena struck me most about Woolman’s focus: the Court’s hard preference for indirect application of the Bill of Rights (as opposed to the rights within it), and the aligned or consequential withering away of the textual purchase of particular rights in the Bill, ie, the increasingly arbitrary link between the nature of the intellectual exercise at issue and the wording of the particular right in the Bill that is ostensibly being analysed in such an exercise … Had the criminal rights jurisprudence of the Court seen a history of fostering a self-conscious discipline of affording to particular rights their doctrinal autonomy, the apparent ease with which the Court appears to forge ahead on its path toward doctrinal agnosticism would no doubt have been absent, or at least, such slights would have sounded a more jarring note.’)

\textsuperscript{48} Barkhuisen v Napier 2007 7 BCLR 691 (CC); Masiya v Director of Public Prosecutions 2007 5 SA 30 (CC), 2007 8 BCLR 827 (CC); \textit{NM v Smith} 2007 5 SA 250 (CC), 2007 7 BCLR 751 (CC).

\textsuperscript{49} Woolman ‘Amazing’ (n 1 above) 763.
Moreover, evidence exists for the proposition that the Constitutional Court’s failure to set out the structure for Bill of Rights arguments (and that embraces arguments about the nature of direct application and indirect application) has the unfortunate consequence of leading perfectly competent High Court judges to make quite egregious mistakes when undertaking constitutional analysis and to avoid direct application of a constitutional right if possible.

For example, in *Hlophe v Constitutional Court*, the majority of the High Court panel of five judges found that a panoply of Hlophe JP’s constitutional rights had been violated.50 The problem with both the applicant’s papers and the majority's judgment is that they identify no underlying cause of action that was subject to challenge on constitutional grounds. In the normal course of constitutional litigation, the applicant will identify a rule of law that, as currently construed, violates one or more constitutional rights. The challenge is to the law as it currently exists in the hope that a change will vindicate the client’s rights and allow for a more appropriate cause of action to proceed. That did not happen in *Hlophe v Constitutional Court*. Instead, Hlophe asserted that a variety of constitutional rights – equality, dignity, privacy, etc – had been infringed by the actions of the Constitutional Court. He did not claim that various actions in delict needed to be altered to give him the relief he sought, nor did he claim that a statute like the Promotion of Equality and Prevention of Unfair Discrimination Act likewise failed to provide an actionable claim or an appropriate form of relief for damages. More disturbing still is that a majority of the five judge panel found that Hlophe’s rights had been violated *per se* – not by existing law, nor the manner of its enforcement. Again: the assertion of self-standing constitutional claims and remedies is not how things are done here. (*Fose* and subsequent case law has made it clear that the Constitutional Court will be slow to create, new, self-standing, constitutional remedies.) Fortunately, the Supreme Court of Appeal recognised that the *Hlophe* High Court had been wrong both in the manner in which it analysed the claims and, consequently, in the conclusions that it had reached.51

While Professor Michelman and I agree on both this analysis of the High Court’s judgment and with the Supreme Court of Appeal’s correction of course, we may not draw exactly the same conclusions. My conclusion is that many High Court judges fear constitutional analysis or do not know how to handle a constitutional matter. I am further inclined to attribute responsibility for this ongoing fear to a Constitutional Court that has not made patently clear the manner in which constitutional claims ought to be handled. (That is an argument about thinness from which Professor Michelman will likely demur.) Moreover, I tend to believe South African constitutional lawyers and scholars would recognise – and decry – the pattern I describe.

50 *Hlophe v Constitutional Court* (08/22932) 2008 ZAGPHC 289.
51 *Langa v Hlophe* (697/08) 2009 ZASCA 36.
How good a witness for the case against thinness is Professor Michelman? It’s hard to know. His writing and his correspondence recognise that the argument against thinness is ‘formidable’ and that the Constitutional Court’s collective inclination [my locution, manifestly not Professor Michelman’s preferred description] towards minimalism is of ‘dubious’ normative value at this juncture of South Africa’s constitutional history. At other times, Professor Michelman appears reluctant to attribute anything other than ‘occasional opacity’ to the Constitutional Court. I am not certain that much can be made of these different locutions.

5 How best to read the Constitutional Court’s application jurisprudence

5.1 Closing the circle

As I had initially hoped, this reply reflects a fairly charitable attempt to maximise areas of agreement and sharpen the small, but remaining, areas of disagreement between Professor Michelman and myself. More importantly, this exercise in aggressive learning has improved our understanding of the Constitutional Court – and how best to read its texts, and the underlying text, the Constitution.

Some open questions still remain. To a large extent they turn on questions that first animated this exchange: how best to understand the text of the Constitution with regard to issues of application and how best to understand the application jurisprudence of the Constitutional Court.

5.2 Can you teach your application theory in the classroom? Making sense of Khumalo and the post-Khumalo jurisprudence

Here’s an experiment the reader can do in the classroom. Teach a roomful of students Du Plessis v De Klerk and explain how the majority reached its conclusion about the limited reach of the Bill of Rights under the interim Constitution. Show them the text of interim Constitution sections 7(1), 7(2) and 35 and help them to understand why Kentridge J concluded that the Bill of Rights under the interim Constitution was, as all constitutions have traditionally been, limited to regulating relationships between the state and individuals.

The next time that you see them, show them the text of sections 8(1), 8(2), 8(3) and 39(2) of the final Constitution. Ask them to apply these

52 Michelman (n 1 above) 1.
53 E-mail from Michelman on 29 December 2009.
54 As above.
sections to the same facts as those facts at issue in *Du Plessis v De Klerk*. Do not assign *Khumalo v Holomisa* until you have conducted this part of the experiment. Some readers may not be *au fait* with either matter. The two cases tracked largely identical fact patterns and both cases began as defamation cases that turned into constitutional matters when the defence challenged the current common law rules of defamation as violations of the right to freedom of expression.

Here is the result – identical across three institutions: Columbia Law School, the University of Pretoria and the University of the Witwatersrand. Since you have explained that the absence of the words ‘the judiciary’ and ‘applies to all law’ in section 7(1) of the interim Constitution led Kentridge J to conclude that the absence of these phrases insulated common law disputes between private parties from the direct application of the substantive provisions of the Bill of Rights, they will reach the following conclusions about (direct) application (of the substantive provisions of the Bill of Rights) under the final Constitution: The inclusion of ‘binds ... the judiciary’ in section 8(1) of the final Constitution will result in some bright student concluding that the drafters of the final Constitution meant to reverse Kentridge J’s conclusion about the extension of section 7(1) of the interim Constitution: The substantive provisions of the Bill of Rights, will, where appropriate, apply to appropriate common law disputes between private parties. Another bright student will remark that the phrase ‘applies to all law’ in section 8(1) of the final Constitution means just that, namely, that the substantive provisions of the Bill of Rights will apply to all law – not just statutes – and that law ‘embraces’ common law rules, subordinate legislation, international covenants, municipal by-laws and rules of customary law.

It comes as something of a shock, then, when you reveal – or when they do the reading – that O'Regan J did not interpret section 8(1) of the final Constitution in that fashion in *Khumalo v Holomisa*. Instead, you are obliged to explain that Justice O'Regan was concerned that if she interpreted section 8(1) in the obvious or natural manner pressed by counsel, then she would struggle to give meaning – or a purpose – to sections 8(2) and 8(3). Okay. How then to reconstruct section 8 and section 39(2) for your students in a manner that coheres with the gossamer thin five paragraphs of *Khumalo v Holomisa*?

Professor Michelman does not answer that question. What he does, instead, is supply his own descriptive account of what the Constitutional Court in fact appears to be doing with these sections, along with an explanation of why that might strike the Court as the path of least resistance. The question, however, remains. Where does that realist-style treatment leave us who bear the responsibility of teaching students, practitioners and jurists what the

55 See Michelman (n 1 above) 8 - 9, 37 - 40.
apposite sections of 8 and 39(2) mean? Professor Michelman offers two responses: (1) First, it is the law of the land and the cases spell that out; (2) second, it is not impossible to reconstruct what the Court is doing in section 39(2) cases (that might best have been fit under sections 8(2) and 8(3)).

Professor Michelman is correct on both accounts. But I am going to emphasise the words ‘not impossible’. Professor Michelman, as a rule, teaches students at Harvard Law School who are in their seventh (or greater year) of tertiary education at one of the world’s greatest universities. I taught for several years as a lecturer, at my alma mater, Columbia Law School, students in their seventh (or greater year) of tertiary education. For such students, the nuances and the contradictions of the Constitutional Court’s jurisprudence are well within their intellectual grasp.

Let me be unabashedly elitist about the inputs and the outcomes. Current students at the LLB and LLM level in South Africa quite often have a difficult time reading the most simplified (dumb-downed) accounts of the Constitutional Court’s doctrines. Reading and understanding actual cases often proves extremely challenging.

But what goes for students proves equally true for practitioners and jurists in South Africa. As I have already noted, a significant percentage of practitioners and jurists have some difficulty understanding application issues and the structuring of standard bill of rights arguments. It is not, as with our students, a matter of incapacity. In large part, many have not been exposed to constitutional law as a field of study – and so feel more comfortable with more common forms of statutory interpretation or common law analysis (thus the familiarity afforded by section 39(2)). The Constitutional Court bears some responsibility here for not making the structure of Bill of Rights analysis crystalline clear in a jurisdiction that has no extended history of such analysis.

In response to Professor Michelman’s question as to why the structure of application analysis should matter, I have suggested two basic responses. First, naturalness, textual plausibility and non-surplusage favour my account. Second, the often self-defeating effort to teach students, practitioners and jurists how the Constitutional Court constructs Bill of Rights analysis again favours a straightforward exegesis of the text as to how meaning and purpose ought to be allocated amongst the four clauses that have a bearing on the application of the Bill of Rights.

Professor Michelman does not spend all that much time explaining his view of the extension of sections 8(1) and 8(2) of the final Constitution. His text suggests that section 8(1) applies to statutes – whether they govern disputes between the state and a person, or disputes between natural and/or juristic persons. In Professor Michelman’s view, nothing much turns on whether disputes between private parties governed by the common law (or
customary law) should not be governed by section 8(1) or, even more appropriately post-Khumalo, by section 8(2) of the final Constitution.

Now why should that be so? It is certainly not what the text says. Nor does it make any sense of the distinction between the text of the interim Constitution and the text of the final Constitution. Why does ‘applies to all law’ not mean what it says? Why does the binding of ‘the judiciary’ not overcome Kentridge J’s objections in Du Plessis? Should not Kentridge J’s reading of the interim Constitution lead us to the conclusion that common law disputes are subject to the direct application of the substantive provisions of the Bill of Rights in terms of section 8(1) of the final Constitution? Professor Michelman and I might well agree that the drafting of sections 8 and 39 of the final Constitution is so shoddy that some ‘surplusage’ is inevitable. But that hardly justifies the absence of an attempt to give sections 8(1), 8(2), 8(3) and 39(2) distinct and independent roles – after all, the non-redundancy requirement laid out by Justice O’Regan in Khumalo is what obliged her to give sections 8(2) and 8(3) the roles that she assigned to them.

As imperfect as it may be, my good faith reconstruction of Khumalo does greater justice to the text and the non-redundancy requirement laid out by Justice O’Regan in Khumalo. That good faith reconstruction takes account of such banal questions as ‘Why is section 8 of the final Constitution called “application”? and ‘Why is section 39 called “interpretation”? 56

---

56 And what of sec 8(2) of the Constitution? On Frank’s account, everything anyone could ever want to do under sec 8(2) can be done under sec 39(2) of the Constitution. When would one rely on sec 8(2) to govern common law disputes between individuals and when would sec 39(2) kick in? Why are we not to take the non-redundancy requirement laid out by Justice O’Regan in Khumalo seriously? Or how could you teach Khumalo in class, and then tell your students to ignore its reasoning entirely. That genuine trouble regarding the interpretation of sec 8 and sec 39(2) remains is evident from extra-curial remarks made by Deputy Chief Justice Moseneke in an article in the Stell LR: ‘In Khumalo v Holomisa, which was decided under the new Constitution, the Court refused the invitation to overrule Du Plessis v De Klerk. It nonetheless decided that the right to freedom of expression is of direct horizontal application, even when the invasion of the right could have been occasioned by a person other than the state or one of its organs.’ D Moseneke ‘Transformative constitutionalism: Its implications for the law of contract’ (2009) Stell LR 1 8. The Deputy Chief Justice’s footnote is truly disconcerting. The final Constitution expanded expressly the direct application, in terms of sec 8, of the substantive provisions of the Bill of Rights to cover disputes between private parties governed by common law. The interim Constitution, by leaving out such phrases as ‘binding of the judiciary’ and ‘all law’ in sec 7(1) of the interim Constitution, was read restrictively by a majority of the Du Plessis Court to mean that the interim Constitution’s Bill of Rights did not apply directly to common law disputes between private parties. A unanimous Constitutional Court in Khumalo, on law and facts identical in all meaningful ways to the law and facts at issue in Du Plessis, held, unequivocally, that sec 8 of the final Constitution does what sec 7 of the interim Constitution did not: namely, in terms of secs 8(2) and 8(3), rules of common law governing disputes between private parties are subject to the direct application of the specific substantive provisions of the Bill of Rights (where appropriate). Several questions arise: (1) Can the silence of the Court about the status of Du Plessis be read to mean that the central holding of Khumalo did
Professor Michelman does spend enormous energy prying apart my apparent collapse of the various possible relationships between sections 39(2) and 172 of the final Constitution. Our application jurisprudence is certainly the better for his efforts.

What I cannot accept, however, is the impression left by Professor Michelman that it matters not as to whether the Bill of Rights applies directly or indirectly as long as the Court generates an outcome that one can, through extensive reconstruction, explain. A text – and more than a handful of judgments – that cannot be taught in a rigorous way to second-year LLBs is an explanation that does not do the text or our students justice.

not overturn the central holding of Du Plessis? No. (2) Can Du Plessis still be good law when the constitutional provisions upon which it was grounded have been completely altered? No. (3) Does not sec 8(2) expressly reject the claim in Du Plessis that the Bill of Rights does not apply directly to private disputes governed by common law? Yes. (4) Did a unanimous Constitutional Court in Khumalo not expressly recognise that the substantive provisions of the Bill of Rights of the final Constitution applies directly to all law and private disputes, where appropriate, through secs 8(2) and 8(3)? Yes. (5) Acknowledging the force of Frank’s claim that it does not matter whether common law disputes between private parties are engaged by sec 8(2) or sec 39(2), is it still correct, as a matter of law, to say that Du Plessis was not overturned by Khumalo read with the text of the final Constitution? No. (6) Should we be satisfied with the Deputy Chief Justice’s contention that the Constitutional Court’s extant approach to the development of the common law remains the same as it did under the interim Constitution? No. The Deputy Chief Justice’s remarks are that much more disconcerting given that the Technical Committee charged with drafting the Bill of Rights of the final Constitution expressly intended for secs 8(1) and 8(2) to capture all matters involving direct application of a substantive provision of the Bill of Rights (including, specifically, the direct application of substantive provisions to disputes between private parties governed by the common law) and that sec 39 was intended to be exactly what it says – a guide to interpretation of all law in light of the spirit, purport and objects of the Bill of Rights. See C Rautenbauch ‘Remarks on the intention of the Technical Committee regarding the extension of sec 8 and sec 39’ (3 March 2011) (on file with author). See also S v Makwanyane 1995 3 SA 391 (CC), 1995 6 BCLR 665 (CC), 1995 2 SACR 1 (CC) (Chaskalson P) (on the value of travaux préparatoires); S Woolman ‘A curious extra-curial footnote: Deputy Chief Justice Moseneke’s aside and the tenacious hold of state action doctrines across constitutional jurisdictions’ Presentation at South African Institute of Advanced Constitutional, Public, Human Rights and International Law. (The paper suggests that the Deputy Chief Justice’s remarks are consistent with ‘state action’ doctrines adopted in Canada, Germany, Australia, the United Kingdom and the United States – and the desuetude into which academic arguments about direct application often fall. The paper’s first hypothesis is that judges across jurisdictions are more comfortable with indirect application than they are with potentially upsetting the common law apple cart through direct application. It appears to matter not what the actual text of any given Constitution says. The second hypothesis turns on the indisputable fact that judges have started to speak globally about constitutional law. This global conversation may reinforce existing practices on the bench rather than disentrench them. That is, the global constitutional conversation may justify existing practices instead of inviting novel approaches.)
6 Conclusion

'This is not an argument!' 'Yes it is.' 'No it isn't'. 'Yes it is.'
'But I paid for an argument.' 'No you didn't.' 'Yes I did'. 'No you didn't.'
Monty Python's Flying Circus, Five minute argument

I leave it to Professor Michelman (in his rejoinder) and the reader to decide whether I have largely closed the circle between us, provided proper amplification of Davidson’s principle of charity, used sufficiently compelling evidence from the primary and secondary literature for my case for clarity, and made any progress in the ongoing debate about how we ought to reconstruct our application doctrine.

But one last word – from Donald Davidson – on the nature of this extended kibitz with Professor Michelman: ‘[M]inimising disagreement or maximising agreement is a confused ideal. The aim of interpretation is not agreement but understanding.’ In short, the very possibility of the kind of extended kibitz that Professor Michelman and I have enjoyed is contingent upon ‘endless true beliefs’ about the Constitutional Court. To the extent that we still part paths during this colloquy, our remaining differences are little more than subtle, bubble comments regarding the work of South Africa’s highest court.57

---

57 As this piece went to bed, the Constitutional Court held that a private party in a dispute with another private party governed by common law – a trust seeking to evict learners from a public school so that it might use the property more profitably – was subject to the formal terms of the Bill of Rights through sec 8(2) and the substantive terms of Bill of Rights through sec 29(1)(Right to basic education). See Governing Body of the Juma Musjid Primary School v Ahmed Asruff Essay NO CCT29/10 [2011] ZACC 13. The judgment expressly confirms the holding in Khumalo v Holomisa regarding the role of sec 8(2) in determining the direct application of specific provisions of the Bill of Rights to disputes between private parties governed by common law. It is now clear that Deputy Chief Justice Moseneke’s extra-curial remarks no longer reflect the state of the law, and, it would appear, the Deputy Chief Justice’s position on the matter.
CHAPTER 18

OLD KIBITZES NEVER DIE: A REJOINDER TO STU WOOLMAN

Frank I Michelman*

1 Introduction

Thanks to Professor Woolman for his gracious invitation to me to add a few thoughts of my own to his typically spirited, probing, and ground-gaining kibitz instalment.

As his instalment shows, and as the following brief notes will confirm, there is no contradiction (at any rate, none intended) between what I wrote in ‘Charity’ and most of what Stu Woolman has to say here. As will also appear, some questions do remain open between us, even after Stu’s careful accommodation to the limited scope of my claims in ‘Charity’ and my reiterations, in what follows, of those limitations.

2 On the principle of charity

As Professor Woolman reports, I had tried carefully to restrict my appeal to Davidson in ‘Charity’. It seems, however, that I may have gotten

---

* Walmsley Professor of Law, Harvard Law School.

1 F Michelman ‘On the uses of interpretive charity: Notes on application, avoidance, equality and objective unconstitutionality from the 2007 term of the Constitutional Court’ (2008) 1 Constitutional Court Review 1.

2 As a reminder, I wrote as follows: The philosophical underpinnings of the charity principle are far afield from our concerns here ... All that matters from that department is the motivation for adherence to the principle ... The aim of interpretative charity is not generosity toward others, or anything like that. It is not to pay homage, deference, or respect to our interlocutors, or to avoid giving offense ... [Rather, the] aim is ... aggressively to learn what there is to be learnt from puzzles the interlocutors pose to us, by assuming there is method in their madness and doing our best to ferret that out, using everything else we know or can guess (in part from their likeness and kinship to us) about where they are coming from. ‘To see too much unreason on the part of others’, Davidson says, is ‘to undermine our ability to understand what it is they are so unreasonable about’. It is to risk missing issues that might merit our consideration.
too clever by half, and I welcome and commend Professor Woolman’s expert cleanup of whatever misunderstandings I may have sown about Davidson’s philosophical message. Since I do not take Stu to be doubting the accuracy (as far as it went) of my restricted take on Davidson, or even my suggestion that conscious recollection of Davidson’s ‘optimisation’ requirement\(^3\) might sometimes materially aid us in our work as legal commentators and since I, for my part, have absorbed unforgettably the lesson to resist future temptations to wikipedish philosophical conceits, there remains, I believe, no issue between us on the ‘Charity’ point.

3  **On the explanatory sufficiency of the Constitutional Court’s judgments: Thinness**

Stu summons, on this point, an impressive body of supportive testimony, which (I say it again), ‘the Court would do well to consult and ponder’.\(^4\) My claim in ‘Charity’ was not to the contrary. It was that (i) inadequacy of explanation is one thing; (ii) erroneous treatment or disposition of a case is another thing; (iii) a judicial ‘flight from substance’ is still another thing; and (iv) the decisions in two of the cases that Stu had especially picked out for criticism \(NM^5\) and \(Masiya^6\) do not support a complaint on the second and third of those grounds. As I wrote near the beginning, to summarise the claims I would be making:

The controlling opinions in these cases are indeed, as Woolman says, ‘thinly reasoned’, if by that we mean they are in some respects insufficiently explained. It is, however, another question whether these cases have been wrongly or irresponsibly

---

\(^3\) On the explanatory sufficiency of the Constitutional Court’s judgments: Thinness

\(^4\) See Michelman (n 1 above) 4 - 5 (citations omitted).

\(^5\) Masiya v Director of Public Prosecutions 2007 5 SA 30 (CC), 2007 8 BCLR 927 (CC).

Michelman (n 1 above) 4.5 (citations omitted).

See Michelman (n 1 above) 4 (quoting Davidson *Inquiries into truth and interpretation* (1984) 101): Davidson ... summarises as follows: ‘We make maximum sense of the words and thoughts of others when we interpret in a way that optimises agreement’. Davidson meant ‘optimise’ as between thinking that the other must be holding to beliefs (and, relatedly, aims) that differ drastically from our own (else he couldn’t have said what he did), and thinking that we must not have heard him right the first time.

\(^6\) Michelman (n 1 above) 4.

\(NM v Smith\) 2007 5 SA 250 (CC), 2007 7 BCLR 751 (CC).
managed, as measured by reasonably discoverable, valid considerations of law and legal administration.

In particular, I shall be questioning Woolman’s diagnosis from these cases of ‘a court uncomfortable with the direct application of the specific substantive provisions of the Bill of Rights’ and ‘in full flight from any meaningful engagement with chapter 2 of the Constitution’. Whether a wider survey of the jurisprudence would warrant an over-all diagnosis of an excessive flight from substance is a question on which I hazard no judgment here. All I say here is that *NM* and *Masiya* do not, to my eye, support the diagnosis, nor is that, in my view, the best way for us to regard these cases. One feature common to both is the Constitutional Court’s seeming gravitation to its inherent power to develop the common law in terms of Constitution sections 173 and 39(2) as opposed to its judicial review power in terms of sections 8 and 172(1) B when undertaking modification of common law rules under pressure from the Bill of Rights. Woolman believes the Court moves too freely to the inherent power. He associates that tendency, as symptom or cause (or both), with excessive flight from substance. I aim to raise a doubt about any such connection.7

In that set of claims, all that may still remain actively contentious between Woolman and Michelman is my denial of a correlation between the Constitutional Court’s (undoubted) gravitation to section 39(2) and the total sum of the expositions and elucidations so far received from the Court of the concrete, substantive requirements for which the several rights-granting clauses in the Bill of Rights are meant to speak.8 I return soon to that matter, but I need first to say a bit more on the general idea of a flight from substance.

### 4 On flight from substance and judicial pragmatism

Given that I have demurred on the question of ‘an overall diagnosis of a flight from substance’ and that Stu (for now) withdraws his claim to that effect,9 there is no active contention between us on that point. There may, however, be some shades of difference between us worth noticing. Stu marshals a set of observations by himself and other scholars regarding the Constitutional Court’s

---

7 Michelman ‘Charity’ (n 1 above) 2 (citations omitted). I added, in a footnote, that I would not be addressing another arguable symptom or cause of flight that I know to be of concern to Professor Woolman, which is ‘the Court’s readiness on some occasions to move to the ‘justification phase of a Bill of Rights case either without having decided the question of a substantive infringement, … or, perhaps, having found an infringement on the basis of a merely ‘notional’ reading of the right in question. Michelman ‘Charity’ (n 1 above) 2 (citations omitted).

8 That Stu still maintains some attachment to the view I deny appears from his pointing to the Constitutional Court’s readiness to treat constitutional challenges to common-law norms under sec 39(2) (as opposed to sec 8(2)) as one factor among several that might, in combination, support a revival of the flight from substance account. See S Woolman ‘Between charity and clarity: Kibitzing with Frank Michelman on how to best read the Constitutional Court’ in S Woolman & D Bilchitz *Is this seat taken? Conversations at the Bar, the bench and the academy about the South African Constitution* (2012) 391.

9 See Woolman (n 8 above) 393 - 394.
performance that might herald a gathering opinion that a ‘flight’ claim is indeed in order. Included are observations concerning ‘the absence of a minimum core in socio-economic rights, rule aversion in criminal constitutional law decisions, the abuse of the reading down of statutes to avoid engaging direct constitutional claims, the substitution of settlements for norm generating remedies, and the notional approach to constitutional rights interpretation’.10

As will easily be noticed, at issue in these controversies are several dimensions of the postures and strategies of courts engaged in the work of constitutional review which draw attention these days from constitutional-legal theorists and commentators: the relative merits and uses of standards as opposed to rules, of balances as opposed to categories, of a ‘dialogical’ as opposed to an authoritarian posture of courts in relation to governments and parliaments, and of the pacing and calibration of judicial remedies when violations are found. I doubt that all of these considerations are easily or informatively reducible to a single measure of a judicial flight from (or a rush to) substance.11

There can be no doubt, though, that all of them are connected to a wide-ranging debate over ‘minimalism’ and ‘pragmatism’ in constitutional adjudication. Stu Woolman has culled from our correspondence a general appraisal from me, to date, of the Constitutional Court’s achievement on this adjudicative strategic terrain, which I think worth repeating for whatever it may disclose about a possible, remaining difference between us:

Stu ... You make good and fair points. They go to the debate over pragmatism and minimalism. I agree that a general policy of minimalism at the Constitutional Court would be dubious for South Africa at this point. So, while I would not myself, on the whole record, characterise the Constitutional Court's performance as striking a badly wrong balance, I certainly see the strength and importance of your view.12

Here, the possible issue between us goes beyond a choice of preferred phraseology. I believe that judicious (not knee-jerk or routinised) applications of minimalism will always have their uses, including in South Africa now – a generalisation from which I would expect no disagreement by Professor Woolman. Where we may come somewhat apart is that I have not perceived the Constitutional Court to be veering, in general, in an excessively minimalist direction. As testimony to my continuing receptiveness to broadly pragmatic-strategic considerations for the South African judiciary, I hereby affirm my approving references, in ‘Charity’, to minimalist-tending moments

10 Woolman (n 8 above) 401 - 410.11 See generally, on this point, M Tushnet Weak courts, strong rights: Judicial review and social welfare rights in comparative constitutional law (2008).12 Email from Michelman on 16 March 2010 (on file with author) (responding to Woolman (n 8 above).
(or moments of ‘judicious avoidance’) in the decisions I there discuss, and to Roux’s analysis in his ‘Principle and Pragmatism on the Constitutional Court of South Africa’.

5 On ‘application’ and teachability

Back, now, to my denial of a correlation between the Constitutional Court’s ready resort to section 39(2) and any tendency toward flight from substance that Stu Woolman or others may be detecting in the Court’s work as a whole. I stand, for now, on my case for denial of any such correlation, as made in parts 3.11 and 4.4 of ‘Charity’. Suppose my case is found persuasive. The question then arises: What else of practical import remains, that might be at stake in Professor Woolman’s disagreement with the Constitutional Court over the right way to knit together the Constitution’s clauses bearing on the application of the Bill of Rights to the substantive content of the common law?

A part (certainly not the whole!) of Stu Woolman’s answer comes in terms of a concern about the transparency and teachability of constitutional-legal doctrine. To that concern, it would be true, but not quite adequate, for me to respond that there is nothing especially unteachable about my ‘realist’ account of the doctrine in at ‘Charity’. You simply tell your students: ‘In apparent opposition to what some able commentators maintain would be a superior parsing of the relevant constitutional texts, the plain fact is that South African courts, starting with the Constitutional Court, do inquire into the common law’s compatibility with the several, specific, rights-granting clauses in the Bill of Rights (and not just into compatibility with some residual ‘spirit’ of the Bill of Rights) while ostensibly performing their obligations, in terms of section 39(2), to develop the common law in a manner harmonious with the Bill of Rights.’ That account of the doctrine may, from any given teacher’s standpoint, report a truth that is unfortunate, but unfortunate does not mean unteachable (a fact for which every teacher, everywhere, every day, must be unutterably thankful).

Well, thankful yes and thankful no. Unfortunate is unfortunate, teachable or not. It would be unfortunate, indeed, to have to teach a rising generation of South African lawyers that close and tough reasoning with the constitutional enactments as written, along with the rest of what goes with a trained and disciplined, lawyer-like regard for constitutional laws as presumably carefully

13 See Michelman (n 1 above) 34 - 35, 47 - 48, 53 - 57.
15 I review this disagreement and consider the possible stakes in Michelman ‘Charity’ (n 1 above) 8 - 9, 35 - 40.
16 Woolman (n 8 above) 415 (typescript).
17 Michelman (n 1 above) 8 - 9.
wrought by focused and competent drafters, are all dispensable as long as the results come out approximately okay. On that point, I am with Stu Woolman. And I can join him, too, in wishing the Constitutional Court not to be too ‘thin’ in its own accountings of its constructions not just of the clauses on application but of the Constitution's text in general.

With all that said, a difference (whose origins Stu describes in the opening of his ‘kibitz’) still remains between us two on the application terrain. That difference consists in my persisting doubt about how great an exegetical advantage, if any, Stu’s proposed parsing of the texts on application holds over that of the Constitutional Court.19

---

1 Introduction

This paper reflects on two instances of contested openness occurring in the course of the recent saga involving Hlophe JP and the judges of the Constitutional Court. While both gripping and significant, that broad and significant story itself is not the focus here. Instead, this work will examine two specific events within that larger narrative. The first is the publication by the Constitutional Court of the fact that the judges of that Court were laying a complaint against Hlophe JP with the Judicial Services Commission (JSC) and the upholding of that decision in the courts. The second is the pair of decisions by the JSC to hold closed hearings on the Hlophe matter and the reversal of both of those decisions in court. In respect of the first event, the Constitutional Court judges’ publication and media statement was itself the subject of two judicial decisions, one in the South Gauteng High Court and one in the Supreme Court of Appeal (SCA). In respect of the second event, whether the JSC hearings would be open or closed was the subject of a public submission process initiated and conducted by the JSC as well as two judicial decisions, both going in favour of openness and against the JSC. This paper will cover the expressed reasoning in these events as well as –
This topic is a worthwhile and appropriate one for a colloquium such as the current one. The openness aspects of the Hlophe JP saga were clearly important, but have not yet been examined at depth. For instance, the decisions over whether or not to close the JSC were keenly anticipated in the media, but were framed largely as a for or against decision regarding Hlophe JP.\textsuperscript{4} The issue became the person. Thus, it was the outcome that was reported on by the media and discussed by the commentators rather than the matter of principle. Likewise, the decision to issue the media statement by the Constitutional Court judges as well as Hlophe JP's decisions to counter-complain before the JSC and to sue in the courts have largely been examined as particular fights within a larger battle rather than on their own. This paper aims to begin to redress that inattention.

As an initial argument, I will argue that the final resting points of the law in both instances – that the JSC hearings be open and the determination that the media statement did not violate Hlophe JP's rights – were correct and, further, that both resolutions may best be understood in terms of a concept of judicious transparency. This concept differs significantly from two other types of transparency that were also often deployed and invoked in the Hlophe JP saga, types of transparency which one could term media transparency and public transparency.\textsuperscript{5}

\textsuperscript{3} It is beyond the scope of this paper to consider the transparency aspects of the judicial discipline procedure to be put into place with the Judicial Service Commission Amendment Act 20 of 2008. See D Milo 'Presume openness not secrecy' Mail and Guardian 4 September 2009.

\textsuperscript{4} See eg, ‘Letter to the editor’ by Calamity Jane, Without Prejudice April 2008 4: ‘The only person who might stand to benefit from a secretive hearing of the evidence is that champion of the implausible denial, the Judge-President in question.’

\textsuperscript{5} In their response, Calland & Oxtoby have expressed frustration that I do not in this piece clearly articulate a general theory of transparency and the place where the concept of judicious transparency fits into such a general theory. They note that it may be ‘somewhat unfair to do so’ but ask whether my ‘approach to ‘judicious transparency’ does contribute to a theory of transparency’. Their view is apparently that I might be but they are not sure. They do think that I am breaking ‘new ground’ and have presented a ‘novel’ notion. Unclear as to the meaning of judicious transparency, they also see my piece falling short in that I do not clearly articulate the notion of judicious transparency. As regards the first, in this colloquium presentation I did not set out to provide a general theory of transparency. Instead, I took the opportunity of the colloquium to reflect on the uses of transparency in the Hlophe JP and Constitutional Court saga. As Calland & Oxtoby note, this is ‘welcome’. I took a topic that was at the same time broader and narrower than the one that Calland & Oxtoby seem to wish me to address. It is broader in that it draws not only upon the right of access to information and the global spread of transparency initiatives, but also on the right of freedom of expression and the right of access to court. As noted in Klaaren \textit{SALJ}, these rights have been fashioned by the Constitutional Court into a concept of open justice. In my view, open justice is the appropriate constitutional background against which to evaluate the instances I chose rather than the global spread of access to information initiatives. It is narrower as a reflection on two instances of openness in the context of a particular political, legal and ethical story, the contest between Hlophe JP and the Constitutional Court, a story which may be told and analysed in various ways and, indeed, which may
2  The media statement and its judgment

The following six paragraphs taken from the SCA judgment give the relevant sequence of events prior to the issuing of the media statement by the Constitutional Court as well as the context of its issuing:

[10] During March 2008, the CC heard the Thint/Zuma appeals from [the SCA]. They were of public interest and importance since they concerned the prosecution of a high-ranking politician, Mr Jacob Zuma, on a number of counts. One of the issues related to legal privilege. The CC reserved judgment. It was ultimately delivered after the events that feature in this judgment [...].

[11] Towards the end of that month [Hlophe JP] visited Jafta AJ who concluded that the respondent had attempted to influence him to find in favour of Mr Zuma. Knowing that the respondent intended to visit Nkabinde J, he warned her of the possibility that the respondent might repeat his attempt.

[12] The anticipated visit to Nkabinde J took place on 25 April, and she, too, concluded that the respondent had sought to influence her. At the beginning of May and soon after the court term began Nkabinde J made a report to another appellant and through her the matter was taken up with other members of the court. They met in the absence of two appellants, discussed the subject, and eventually agreed to lodge a complaint of judicial misconduct against the respondent with the JSC based on the information provided by the two Justices. This was done on 30 May.

[13] The gravamen of the complaint was in these terms:

‘A complaint that the Judge-President of the Cape High Court, Judge John Hlophe, has approached some of the judges of the Constitutional Court in an improper attempt to influence this Court's pending judgment in one or more cases is hereby submitted by the judges of this Court to the Judicial Service Commission, as the constitutionally appointed body to deal with complaints of judicial misconduct.’

The document identified the case involved and stressed that there was no suggestion that any litigant was aware of or had instigated the respondent’s
action. It contained further statements about the seriousness of the conduct; the democratic values contained in section 1 of the Constitution; the independence of the judiciary and the prohibition in section 165 of interference with courts; the judicial oath; that attempts to influence a court violates the Constitution and threatens the administration of justice; and that the CC and other courts would not yield to or tolerate attempts to undermine their independence.

[14] A media release in virtually identical terms soon followed, which was sent automatically and electronically to all subscribers to the CC’s information system.

[15] It should be noted at this early stage that (a) the respondent was not apprised of the allegations or their source; (b) he was not asked for his version or comments; (c) he received no effective prior notice of the intention to lodge the complaint; and (d) he was not told of the intention to issue a media statement. The public, too, was not given any detail and was left with nothing more than the knowledge that a complaint with serious implications had been lodged.

Here endeth this excerpt of the SCA’s judgment. The media statement issued by the judges of the Constitutional Court is given in full in Appendix A of this paper. In outline, the media statement noted that Hlophe had ‘approached some of the judges of the Constitutional Court in an improper attempt to influence this Court’s pending judgment’ and that the Court had referred a complaint to the JSC. The statement further noted that the judges of the Court ‘view conduct of this nature in a very serious light’. The Court’s final paragraph passed at least implicit judgment by stating: ‘This court … will not yield to or tolerate unconstitutional, illegal and inappropriate attempts to undermine their independence or impartiality.’

The first court to judicially consider this sequence of events was the South Gauteng High Court, acting upon a complaint laid by Hlophe JP against the conduct of the judges of the Constitutional Court. In a somewhat convoluted judgment, that court decided in favour of Hlophe JP and against the judges of the Constitutional Court. The Court was, however, divided by three judges to two, with the two minority judges writing separate dissents. As the Supreme Court of Appeal noted, the High Court panel was ‘unusually though permissibly constituted as a full bench with five judges …’ Further, while it is arguably not relevant, as several

7 The response of Calland & Oxtoby pushes me to engage further with the rationales of these High Court judgments. For a brief instance of such, see J Klaaren & G Penfold ‘Just administrative action’ in S Woolman et al Constitutional law of South Africa (2002) 63-14 fn 4. The point made here is that the SCA did what the High Court did not; it separated out and focused on the precise question of publication of the statement. In so doing, the SCA considered the matter within the context of openness and public justification rather than within the context of procedural fairness understood as fairness to Hlophe JK.
8 Langa v Hlophe (697/08) [2009] ZASCA 36; 2009 4 SA 382 (SCA) (31 March 2009) para 1. This was per authority of Supreme Court Act 59 of 1959 sec 13(1)(a).
commentators did point out, the three majority judges were African; the two dissenters white. The majority reasoned that fairness in the circumstances dictated that an accused judge should be heard when a senior person without personal knowledge of an incident considers laying a complaint regarding a breach of judicial conduct.

Taking the appeal from the South Gauteng High Court, the SCA dismissed the majority reasoning regarding the fairness of Hlophe JP's treatment by noting that there is no principle for an accused to be heard by complainants before a complaint is laid and by characterising the special treatment due to an accused judge by his or her judicial colleagues as a matter of ethics rather than of law enforceable in courts. The SCA then considered in paragraphs 48 to 55 of its judgment the question of whether the media statement was unlawful by publishing the fact that the complaint had been made to the JSC. In the view of the SCA, this question was separate and distinct from the issue of whether Hlophe JP should have been heard before the complaint was made. Further, the SCA recognised that this precise question of publication had not been considered by the High Court, since it viewed the laying of the complaint and the issuing of the media statement as intertwined. The SCA thus distinguished the aspect of openness from that of procedural fairness, reframing the case. Nonetheless, the SCA's consideration of the lawfulness of the publication clearly linked the resolution of that question with the question of the lawfulness of the laying of the complaint. The SCA stated:

Once having found the appellants did not act unlawfully in laying the complaint we can see no basis for finding that they were obliged to keep that secret for the reasons dealt with more fully below. On the contrary there is much to be said for the contrary proposition (bearing in mind the circumstances in which it occurred) that the constitutional imperatives of transparency obliged them to make the fact known.

The SCA then proceeded to quote [presumably] and with apparent approval from the submissions of the Constitutional Court judges as follows:

In the circumstances where the independence of the Constitutional Court had been threatened and the integrity of the administration of justice in South Africa generally, it was considered imperative and appropriate that this be publicly disclosed. Should the facts have emerged at a later stage there would have been a serious risk that the litigants involved in the relevant cases and the
general public would have entertained misgivings about the outcome and the manner in which the decisions were reached. It was especially important that the litigants and the general public were informed of the attempt and that the Constitutional Court had not succumbed to it.

The rationale expressed in this context by the judges of the Constitutional Court and adopted by the Supreme Court of Appeal is significantly circumscribed. Stripped to its bare essentials, the reasoning may be paraphrased as follows:

If word of the attempt to influence us got out, and it was not made clear that the attempt had failed, then the public would doubt our judicial decision-making process but, never fear, we are hereby telling the public that the attempt to influence us did fail and our decision-making processes are just fine and intact.

In my view, this is a rationale based on the Court’s core judicial function. At base, the SCA is agreeing that the Constitutional Court’s ability to judge (indeed, the prime function of the Constitutional Court) was being placed in issue by the fact of the contact/approach. Another relevant implication that one might draw from the statement is to the effect of ‘we think that there is a significant possibility that this word might get out. The Constitutional Court thus did not appear to have total confidence in the informational integrity of its own organisational and decision-making processes, as I shall elaborate on further below.

In the view of the SCA, which has had the final judicial word here, the media statement was not unlawful. While the media statement may have been prima facie unlawful for carrying the implication that Hlophe JP had attempted to improperly influence other judges, a statement may be justified and that justification ‘can be raised validly if the statement was true and for the public benefit’. Indeed, according to the SCA in its elaboration of the law of defamation, ‘[d]isclosure of an allegation of gross misconduct against a judge may in certain circumstances not be for the public benefit but that could hardly be the case if the allegation is true’. Thus, for the SCA, the possibility that the prima facie defamatory allegation in the statement was true thus needed to be considered in deciding whether the publication was unlawful. An assessment of the truth of the allegation had not been done at the High Court level, of course, and the SCA thus overturned the High Court. The appeal succeeded. In this way,

---

14 It perhaps thus ironic that it was the determination that the judges were not exercising a judicial function in making the complaint that decided the lawfulness of that question. See eg Langa v Hlophe (n 9 above) para 47.
15 Langa v Hlophe (n 9 above) para 51.
16 Langa v Hlophe (n 9 above) para 54.
17 The response of Calland & Oxtoby pushes me to link this decision to previous case law on defamation. It would indeed be interesting and worthwhile to do so, but a full analysis of the decision from the viewpoint of defamation case law and of the place of various rationales of openness within such case law would lie beyond the scope of this paper colloquium presentation.
without itself actually examining or determining the truth of the media statement, the SCA was able to determine the statement not to be unlawful.

3 Opening and closing the JSC hearings

Given the mind-bending and attention-sapping twisty and tortuous path of the JSC consideration of the Constitutional Court judges’ complaint (not to mention Hlophe JP), this paper may be forgiven for merely summarising the three decision-making instances regarding the question of opening the JSC’s disciplinary proceedings for Hlophe JP. The first instance occurred in July 2008 when the JSC itself, in a commendable display of openness to public participation, decided to and did invite submissions on the question of whether the disciplinary hearings to be conducted in respect of Hlophe JP were to be closed or open to the public and the media (as for instance are the interviews of candidates for judicial office). This was done after Hlophe JP had made his counter-complaint. A number of persons and organisations made submissions, including the Wits Law School. The second instance occurred after the JSC had decided on 28 March 2009 to close the hearings. On the Saturday before the Wednesday scheduled start of the hearing, the JSC announced its decision: that the proceedings would be closed. As Milo describes it:

This decision prompted an urgent application, the day before the hearing was to start, by most of the major media groups – Avusa, Independent, Mail and Guardian, Media 24 and e-TV – and the Freedom of Expression Institute and the Centre for Applied Legal Studies. The JSC argued that the hearing should be in secret in order to protect the dignity and statute of the office of the chief justice and the deputy chief justice of the Constitutional Court and of the judge-president, all of whom would be required to give evidence at the enquiry.

The JSC also noted that they would give reasons after the hearings. In the High Court, Judge Nigel Willis rejected the JSC’s argument ‘ruling that the dignity and statute of the judiciary as a whole would be enhanced rather than undermined if the hearings were to be held in the open: The

18 D Milo ‘The very soul of justice’ (June 2009) Without Prejudice 12.
19 See eg ‘Submission to the Judicial Service Commission by members of the School of Law of the University of the Witwatersrand, July 2008’. The Wits legal academics argued in part that ‘[i]n a matter that has attracted enormous public interest, open hearings could have an important educational value in allowing the public to understand the issues at stake and the nature of judicial independence. By contrast, closed hearings will limit public knowledge and information, and limit the public’s ability to engage with the issues.’
21 Judicial Service Commission ‘Media statement, 31 March 2009, Justices of the Constitutional Court and Judge-President Hlophe, Reasons for decision not to hold hearing in public’.
The third instance consists of another court application, occasioned by the JSC’s decision on 20 July 2009, despite the then recent High Court decision, to hold a closed preliminary hearing into the complaint against Hlophe JP. In this instance, some of the applicants that went to court on behalf of openness went further than contesting the question of transparency. Some asked for a ruling regarding the substance of the JSC process. Rejecting such a substantive intervention, the Court, per Malan J, gave a decision that the openness as decided in the earlier case needed continued respect.

As has been noted, these decisions on transparency were to some degree mere skirmishes in the larger campaign regarding Hlophe JP and, in the view of some, the transformation of the judiciary itself. Nonetheless, these were important points of conflict, both for the campaign within which they were waged as well as for constitutional and open democracy more generally. After the decision not to proceed with the complaint was announced, Andrew Brown, a perceptive and witty Cape advocate, pushed further the question of openness in a comment in the following terms:

If, as has been suggested, the decision of the JSC has resulted in the undermining of public and professional confidence in the judicial system, then the fault surely must lie squarely at the feet of the members of the JSC, not with any one judge and his alleged indiscretions. That being so, the attentions of the disenchanted should be focused squarely (and I would suggest exclusively) on the weaknesses that are to be found in the JSC’s make-up (its constitution, not its domestic façade) and its processes. If, as has also been suggested sotto voce, the voting on the JSC on this issue was split along racial lines, then we need to establish alterative procedures that could cure the JSC of an apparent racial malady. Is the procedure for the nomination of persons to serve on the JSC adequate? Should the proceedings of the JSC be open to public scrutiny? If voting were open and public, would members of the JSC be more accountable?

4 Analysis

As I shall argue in this section, the final legal determinations – of the SCA that the media statement did not violate Hlophe JP’s rights through publication and of the South Gauteng High Court that the JSC hearings be

---

22 eTV (Pty) Ltd v Judicial Service Commission (n 21 above) 16.
open – were correct as matters of law. It also and significantly seems correct that the July 2009 decision regarding the JSC did not go further than considering the question of openness.

In my view, the correctness of these two final legal determinations may best be understood in terms of a concept of judicious transparency. Judicious transparency is a form of openness ultimately linked to the Constitution’s notion of open justice, as articulated in the Independent Newspapers case.26

In at least two dimensions, judicious transparency may be distinguished from other forms and rationales of transparency. First, judicious transparency is distinct from what might be termed presumptive or media transparency. To paraphrase, the media have almost never met an open hearing or proceeding that they have not liked. Taken for granted is a presumption of openness that is the opposite of considered and nuanced. This is practically a principle of professional qualification in the field. Indeed, media transparency may itself be a version of the second-best argument that Joseph Stiglitz has made on behalf of transparency.27 His argument with respect to globalisation and the place in that of international public legal institutions begins with a recognition that the first best thing at the moment is to change the governance structures of structures like the International Monetary Fund, the World Bank and the World Trade Organisation. He then makes a clearly-defined ‘a half loaf is better than no loaf’ argument:

Short of a fundamental change in their governance, the most important way to ensure that international financial institutions are more responsive to the poor, to the environment, to the broader political and social concerns that I have emphasised is to increase transparency and openness.28

In addition to its differences with media transparency, judicious transparency is also distinct from what can be termed public transparency, though it shares with public transparency many of the goals inherent in the support of constitutional democracy.29 As an instance of public transparency, we can take the quote attributed to Prof Cathi Albertyn in

27 Stiglitz, Calland and others are engaged in research on the global wave of transparency linked to the right of access to information. See eg C Darch & P Underwood Freedom of information and the developing world: The citizen, the state and models of openness (2010). As articulated above (n 6), my topic in this piece has differed from that topic. However, I support Stiglitz et al and their pragmatic espousal of transparency practices in general. Such practices most often are an immediate and effective means by which to exert pressure towards change towards social justice.
29 Discussing only the example of Prof Albertyn’s appeals in the context of the recent JSC interviews of judicial candidates, the response of Calland & Oxtoby argues that the rationale of public education does not exhaust the notion of public transparency. I would agree. The rationale of the Constitutional Court submission to the SCA noted in
Legal Brief in October 2009. This was on the occasion of the JSC’s decision, as is its constitutional duty, to nominate seven candidates for posts as judges of the Constitutional Court from the 22 short-listed names it had previously decided upon. Albertyn is paraphrased as saying the way the JSC currently operated was not transparent and quoted as its operation resulted in ‘us not ending up with the best constitutional lawyers because we are too busy looking at other things’. Upon closer examination, Albertyn’s call resolves into two separate appeals. One is that the criteria for judicial selection be better articulated. This call has been echoed elsewhere, for instance, in a well-crafted editorial by former Constitutional Court researcher Susannah Cowen. The second appeal relates to the type of questioning engaged in by the JSC commissioners. Albertyn perceived, as indeed many members of the public at large do, that some candidates in the JSC got an easy run and some got a hard run. In Albertyn’s view, there should have been both greater consistency and, in particular, greater attention to constitutional conversations. The interviews were like job interviews but should have been conducted at a higher level. In this view, the primary value of the potential openness of the hearings would be their public educational potential.

Distinguishable from both of the above forms of transparency, the concept of judicious transparency is rooted in the constitutional concept of open justice. This concept is itself derived from the rights of freedom of expression, the right of access to information, and the right of access to court. It is further constituted by the direct implementation by judges of the constitutional principle of openness. A practice of judges related to demands for openness, judicious transparency is a particular form of transparency. I would argue that it is not concerned, at least in the first instance, with the individual dignity of any particular judge. Instead, it is its manner of implementation that makes it particularly judicious. As for its substantive value, the substance of that value is indicated by the three rights from which the concept of open justice has been constructed.

The text would constitute another example as would Willis J’s statement regarding the JSC hearings and Andrew Brown’s comment. More generally, the concept of public transparency would seem to be completely included in the concept of open justice.

Legal Brief 6 October 2009.
The JSC did attempt to meet a request for disclosure of the criteria. See ‘Re: Request for Access to Record of Public Body in terms of Section 18(1) of Promotion of Access to Information Act 2 of 2000’ V Masangwana to M Desai (3 April 2009)( listing in a paragraph the ‘wide variety of factors which are taken into account’).

S Cowen ‘What exactly are we looking for in the ideal SA judge?’ Business Day 17 September 2009. See also S Cowen ‘Judicial selection’ (cited in Response by Calland & Oxtoby n 17).

See Open justice and beyond (discussing the appropriate test for limiting open justice).
The response of Calland & Oxtoby spends much of its time arguing that the meaning of the concept of judicious transparency is unclear. They ask: ‘But is it the fact it is courts or other such bodies acting which is important, or the nature of actions?’ It is both. As I noted later in the conclusion: ‘There was no intermediary between the judges and the action that needs to be judged; it was not some official that took the action, the judges themselves did the deed.’
The argument here is not that the lodging of the complaint was either politically wise or ethically necessary (or even correct). Those questions of politics and of ethics are ones that are distinct from the admittedly narrow conception of legal correctness and its rationale in a concept of judicious transparency that I am exploring here. The question of whether the judges of the Constitutional Court were wise to lodge the complaint has been explored in some depth by Roux from a law and politics perspective. Roux concludes that the decision to do so was ‘truly disastrous’ from the perspective of ‘the long-term project of subordinating political power to the rule of law’ and thus that the action of the judges was ‘regrettable’. Roux argues that, while the alleged conduct was serious enough to constitute gross misconduct, the judges should have considered the likelihood that the allegations could be prosecuted to a successful conclusion and the negative implications for the lodging of the complaint ‘for public confidence in the Constitutional Court and the administration of justice in South Africa generally.’ Apart from the political wisdom, while there has been extensive discussion by commentators in the media, a comprehensive analysis of the ethics of the action of the judges of the Constitutional Court remains to be undertaken. The judges of the Supreme Court of Appeal did seem to indicate that, as an ethical as opposed to a legal matter, the Constitutional Court judges’ action might well be judged differently and negatively. The judicial ethics argument providing the foundation to the

Roux notes that he is assessing the action against ‘prudential criteria’. Roux (n 2 above) 9. In constitutional law and politics, prudential doctrines are ones crafted by courts to husband their institutional legitimacy and/or other resources. They thus guide the decision of whether or not to exercise legal judicial power and assume that legal power may be exercised. In other words, as Roux puts it, they are concerned really with the wisdom rather than the correctness of the decision. Roux (n 2 above) 8 (not discussing in any depth the SCA decision).

Roux (n 2 above) 11.

Roux (n 2 above) 8.

Roux (n 2 above) 8.

The response of Calland & Oxtoby does not take up this invitation, although Calland did pronounce on this question at the colloquium at which this paper was initially delivered. He called for the development of judicial ethics and I agreed. And here is perhaps the sole hard distinction, as far as I can tell, between the response of Calland & Oxtoby and my paper. Calland & Oxtoby call into question ‘drawing a distinction between how the complaint is viewed ethically and legally’. In their apparent view, the constitutional basis for argument on both questions means that there is or should be no distinction. There is much to be said for the Calland & Oxtoby position that the resolution of questions of ethics directly and immediately translates into the resolution of questions of law in situations such as the instances of openness I have discussed in this paper. Indeed, judges pride themselves in following ethical conduct and the dangers of positivism. But, as we know in our constitutional order and perhaps know even more deeply following the unrolling of this Hlophe JP Constitutional Court saga, our judges are human. And it would seem too far-reaching a principle to collapse the ethical enquiry into the legal one in these instances. To do so might lose much of the useful room for maneuver, criticism and reconciliation that still exists in the institutions and offices that regulate and guide the conduct of the judiciary.

‘There is considerable merit in the submission that a judge who is minded to lay a complaint against a colleague has special duties that are not shared by lay complainants, for there is an overarching duty upon judges, in whatever they do, to preserve the dignity of the judicial institution. Indeed, the Constitution itself commands all organs of state, which include the judiciary, to assist and protect the independence, impartiality, dignity, accessibility, and effectiveness of the judiciary.’ The duty that is cast upon judges no doubt calls upon them to act with due care and circumspection.
SCA’s suggestion has not yet been furnished, but might well be furnished by one of the perceptive and informed commentators on the place of the judiciary in the South African constitutional democracy, such as Calland. Interestingly, Roux does note that the most feasible alternative course of action open to the Constitutional Court judges was to have called Hlophe JP to for his account to Langa CJ in his capacity as the Chief Justice and the head of the judiciary. To do so would have been to choose an ethical rather than a legal battleground. Yet, it was a legal battleground that was chosen – by both sides – and that did indeed provide the weapons of conflict in both these incidents.

5 Two legal conclusions

Shorn of the political and ethical dimensions (admittedly a difficult and perhaps dangerously positivistic move), I see a correct legal determination of judicious transparency in the issuing of the media statement by Constitutional Court judges for three reasons. First, the subject matter was on an important and delicate public matter, a fact situation at the intersection of partisan politics, the politics of judicial personnel, and the politics of the judiciary within a constitutional democracy. This was a statement by judges about a judge and about judging. Second, this was an instance of judicious transparency since this transparency had to be put into action by individual judges. Their activity constituted a judicial practice. There was no intermediary between the judges and the action that needs to be judged; it was not some official that took the action, the judges themselves did the deed. Thus, it was the judges that were liable in terms of the applicable law of defamation, the private law arena in which the SCA judgment was ultimately fought out. Third, the judiciousness of the transparency practice needs to be examined within its actual institutional context. In particular, an element of context that is important to take into account and is properly taken into account goes beyond the narrow judicial process reasoning given by the SCA. In my view, the credible evidence here that mentions the potential role of agencies of national intelligence must be taken into account and has, by and large, not been highlighted in

\[40\] Roux (n 2 above) 7.

\[41\] The publication of the lodging of the complaint was thus a private law analogue to the direct judicial conduct regulating access to court at issue in the Independent Newspapers case.
Judicious transparency

the discussion of this sequence of events. This possibility goes beyond any leakage of information from the Constitutional Court and entertains rather some interference with court operation. A potential role of the agencies of national intelligence in a matter such as this could effectively constitute a broader attack on judicial independence and the administration of justice and arguably on constitutional democracy. The Court may have been spooked by the spooks. From this angle, the publication was thus an instance of judicious transparency, appropriately not as a judicial function, but rather as a non-judicial decision taken by those entrusted to deciding judicially.

I also see a correct legal determination of judicious transparency in the judicial re-opening of the hearings after decisions to close them were taken by the JSC. Here, it was two High Court judges that served to remind the JSC of the constitutional value of openness, and of the rights of freedom of expression, access to information, and access to courts. These decisions did not and should not have turned on the right to dignity. While relevant, dignity should not be the framing analysis. Indeed, it may be appropriate to recall in conclusion that the SCA emphasised the equal worth of judges with other citizens. Equal but not more equal.

Appendix A: Statement by judges of the Constitutional Court

1 A complaint that the Judge-President of the Cape High Court, Judge John Hlophe, has approached some of the judges of the Constitutional Court in an improper attempt to influence this Court’s pending judgment in one or more cases has been referred by the judges of this Court to the Judicial Services Commission, as the constitutionally appointed body to deal with complaints of judicial misconduct.

2 The complaint relates to the matters of Thint (Pty) Ltd v National Director of Public Prosecutions and Others (OCT 89/07), JG Zuma and Another v National Director of Public Prosecutions and Others (CCT 91/07), Thint Holdings (South Africa) (Pty) Ltd and Another v National Director of Public Prosecutions (OCT 90/07) and JG Zuma v National Director of Public Prosecutions (CCT 92/07). Argument in these matters was heard in March 2008. Judgment was reserved in all four matters. The Court has not yet handed down judgment.

Roux agrees that this threat constitutes an attempt to influence the decision of Judge Nkabinde, but apparently would not go so far as to characterise the threat as broader. For an official assessment of the precarious positioning of intelligence in the contemporary South African constitutional democracy, see Ministerial Review Commission on Intelligence ‘Intelligence in a constitutional democracy: Final Report to the Minister for Intelligence Services, the Honourable Mr Ronnie Kasrils, MP’ (10 September 2008).

Of course, the South Gauteng High Court decisions were jurisdictionally clearly legal reviews of the decisions of a public body, the Judicial Services Commission. Thus, they do not precisely fit within my definition of judicious transparency. Nonetheless, those court decisions do bear certain similarities in that they are about judges judging a judicial commission that itself judges judges. Moreover, the substantive issue at stake in the judicial review was that of transparency, a key concept in open justice and open democracy.
3 We stress that there is no suggestion that any of the litigants in the cases referred to in paragraph 1 were aware of or instigated this action.

4 The judges of this Court view conduct of this nature in a very serious light.

5 South Africa is a democratic state, founded on certain values. These include constitutional supremacy and the rule of law. This is stated in section 1 of our Constitution. The judicial system is an indispensable component of our constitutional democracy.

6 In terms of section 165 of the Constitution the courts are independent and subject only to the Constitution and the law, which they must apply impartially and without fear, favour or prejudice. No person or organ of state may interfere with the functioning of the courts. Organs of state must assist and protect the courts to ensure the independence, impartiality, dignity, accessibility and effectiveness of the Courts.

7 Each judge or acting judge is required by item 6 of schedule 2 of the Constitution, on the assumption of office, to swear an oath or solemnly affirm that she or he will uphold and protect the Constitution and will administer justice to all persons alike without fear, favour or prejudice, in accordance with the Constitution and the law. Other judicial officers or acting judicial officers must swear or affirm in terms of national legislation.

8 Any attempt to influence this or any other Court outside proper court proceedings therefore not only violates the specific provisions of the Constitution regarding the role and function of courts, but also threatens the administration of justice in our country and indeed the democratic nature of the state. Public confidence in the integrity of the courts is of crucial importance for our constitutional democracy and may not be jeopardised.

9 This Court – and indeed all courts in our country – will not yield to or tolerate unconstitutional, illegal and inappropriate attempts to undermine their independence or impartiality. Judges and other judicial officers will continue – to the very best of their ability – to adjudicate all matters before them in accordance with the oath or solemn affirmation they took, guided only by the Constitution and the law.

30 May 2008

JUDGES OF THE CONSTITUTIONAL COURT
In the last ten years there have been huge advances in the right of access to information (ATI). The idea that ‘transparency’ is important for meaningful and effective democratic process has been widely recognised.\(^1\) Many new laws have been passed; there are now around 80 countries with ATI laws of some sort or other. There is a new – and fast-growing – body of experience and evidence. There are many human stories and the community of practice has grown enormously. But the scholarship has not kept pace. The search is on for a stronger theoretical basis for advancing the right of access to information and the arguments in favour of transparency.\(^2\) While it may be somewhat unfair to do so, given that he may well not have set out to make any contribution to the theoretical understanding of transparency, one primary question that can be posed is whether his approach to ‘judicious transparency’ does contribute to a theory of transparency.

Certainly the paper breaks new ground. The notion of ‘judicious transparency’ is an entirely novel one. Klaaren sets out to examine ‘two instances of contested openness’ during the Hlophe/Constitutional Court dispute. These are the Court’s publication of the fact that it was laying a complaint with the Judicial Service Commission (JSC) regarding Hlophe...

---


\(^2\) One of the writers (Calland) is part of a small group of academics who have agreed to collaborate on a global academic research project on the right of access to information, with the intention of responding to the exponential growth in empirical-based understanding of the practice of transparency law and policy, and to contribute to the development of a theory of transparency and access to information.
Chapter 20

JP, and decisions by the JSC to hold closed hearings. Both events were the subject of subsequent court decisions. Klaaren sets out to ‘cover the expressed reasoning in these events’ and the ‘underlying situation’ as publicly available.

Framed to avoid personalising the issues as being ‘a for or against decision regarding Hlophe JP’, the analysis is especially welcome in this respect. As with so much of the current South African political discourse, the public debate of the issue was generally not characterised by either a nuanced or dispassionate examination of the transparency issues, but often personalised and partisan assumptions of positions for or against Hlophe or the judges of the Constitutional Court.

It remains to be determined, however, how far Klaaren’s analysis goes in shedding light on the underlying transparency issues. We suggest that it does not go as far as it might. Take the bulwark of Klaaren’s analysis, the concept of ‘judicious transparency’: What does this term in fact mean? And is it a helpful concept? Klaaren does not really provide answers to these questions; it is unclear exactly how the concept is to be defined and, whilst Klaaren makes it clear that he regards the two legal decisions central to the paper as correct, it is not clearly established why he believes this to be so.

1 What is ‘judicious transparency’?

At first reading, the exact meaning of ‘judicious transparency’, as advanced by Klaaren, is elusive and somewhat enigmatic, and the explanation fragmentary. The use of the word ‘judicious’ is not that of its plain language usage, namely, sensible and prudent, and ‘having or done with good judgment’. But nor does Klaaren mean ‘judicial’. Presumably because the expression ‘judicial transparency’ would imply ‘transparency by judges’ or transparency within the judicial branch of government, he avoids this phrase though it seems reasonably clear that his application of his own nascent principle of ‘judicious transparency’ is closer to ‘judicial’ than it is to the plain language ‘judicious’ because it is primarily concerned with the judicial implementation of another principle, namely that of ‘open justice’.

In turn, the meaning of ‘open justice’ has not been explicitly set out; the meaning of the ‘practice of judges’ being ‘a particular form of transparency’ is unclear (does this mean judicious transparency is limited to decisions taken by judges, and has no application outside that sphere?); and quite what content is derived from the relevant constitutional rights is not clearly established. As to Klaaren’s argument that it is the ‘manner of

implementation’ that makes transparency ‘judicious’, the question may be asked – to what end? If not for the dignity of a particular judge (since this basis is disavowed), then for what? For the dignity of the bench, or of the legal profession in general, or for something else?

Klaaren identifies two respects in which judicious transparency is distinguishable from ‘other forms and rationales of transparency’. Firstly, it may be distinguished from ‘presumptive or media transparency’ which demands ‘a presumption of openness that is the opposite of considered or nuanced’ – that is, he implies, virtually absolute transparency. But Klaaren does not make clear what factors determine the consideration and nuance that would make transparency ‘judicious’ as opposed to ‘presumptive’. It is also not immediately apparent why – though it is clearly implied – judicious transparency is the better form.

The second form of transparency which is distinguished is ‘public transparency’, although the two are said to share ‘many of the goals inherent in the support of constitutional democracy’. To illustrate the concept of public transparency, Klaaren cites Albertyn making two ‘appeals’ in relation to the recent JSC nominations for vacancies on the Constitutional Court. The first, that criteria for judicial selection be better articulated, is clearly related to the issue of transparency – although again, what distinguishes it as ‘public’ transparency is unclear. However, it is less clear how the second appeal – relating to the type of questioning by JSC commissioners – is related to transparency. It might be said that the issue is not so much an openness problem – in that the hearings are open and people are able to observe that there are problems with the questions being asked – as a procedural issue. How does remedying the problems with what questions candidates are asked link to issues of transparency? Klaaren does remark – presumably in relation to the JSC’s hearings into Hlophe rather than the judges’ interviews – that ‘the primary value of the potential openness of the hearings would be their public educational value’. If this is the only rationale of for ‘public transparency’, it seems a rather limited – although not unimportant in itself – claim for transparency. Moreover, the scope of the transparency it entails is not clear.

Klaaren defines judicious transparency as ‘a form of openness ultimately linked to the Constitution’s notion of open justice, as articulated in the Independent Newspapers case’. Although Klaaren does not expressly articulate it, this must mean something along the following lines: In Independent Newspapers, the Constitutional Court was faced by a claim that the media and the general public had a constitutional right of access to court proceedings. Writing for a majority of the Court, Moseneke DCJ identified a ‘cluster’ or ‘umbrella’ of related rights, comprising the rights of

freedom of expression and the right to open judicial proceedings (including rights of both public trial and access to court), which he defined as ‘the right to open justice’.\(^5\) Moseneke DCJ identified this systemic requirement of openness as being grounded in founding values of the Constitution, ‘which enjoin our society to establish democratic government under the sway of constitutional supremacy and the rule of law’, in order to ‘ensure transparency, accountability and responsiveness in the way courts and all organs of state function’.\(^6\)

The right to open justice informs ‘the media’s right to gain access to, observe and report on the administration of justice’, but Moseneke DCJ also noted that the cluster of rights which constitute the principle of open justice ‘derive from the Bill of Rights and that important as these rights are … they are not absolute’.\(^7\) Moseneke DCJ pointed out further that it is not unusual for democratic societies to limit open court hearings to protect fair trial rights or the rights of vulnerable groups.\(^8\)

Despite, or perhaps because of, Klaaren’s passing mention of the judgment, the reader is not exposed to the line of thought that flows from the development of the idea of open justice in the \textit{Independent Newspapers} case. Klaaren does not explain exactly what this notion of open justice is, and the precise way in which it is furthered by judicious transparency, or how the two principles are connected. One senses that Klaaren is onto something but, tantalisingly, the reader is left to his or her own devices in making sense of the gaps in the line of argument (which may not be an entirely bad thing).

Whilst Klaaren identifies the concept of open justice as deriving from constitutional rights of freedom of expression, access to information and access to court, he does not explain what specifically is derived from these rights, and thus what the precise content of this principle of transparency is. This complaint is not mere pedantry, for the lack of a clearly developed argument makes a clear understanding of the concept of judicious transparency difficult. For example, the \textit{Independent Newspapers} judgment deals specifically with the right of media coverage of and access to the administration of justice as flowing from the principle of open justice. At face value, this appears as if it would be relevant to the concept of ‘media transparency’ also discussed by Klaaren, yet the principle of open justice is expressly linked to, and used to define, judicious transparency. A further issue is that, whilst the Court identifies possible limitations on the rights comprising open justice in the context of open court hearings, Klaaren does not expressly consider how and to what extent limitations may be

\(^5\) \textit{Independent Newspapers} (n 6 above) para 39.

\(^6\) \textit{Independent Newspapers} (n 6 above) para 40.

\(^7\) \textit{Independent Newspapers} (n 6 above) paras 40 & 44.

\(^8\) \textit{Independent Newspapers} (n 6 above) para 44.
imposed in the case of judicious transparency (although he evidently does not see judicious transparency as being unlimited transparency).

Indeed, the explanation becomes increasingly vague, with one of the other sources for judicious transparency being identified as ‘the direct implementation by judges of the constitutional principle of openness’. How this constitutes judicious transparency is again rather obscure, particularly since, as already discussed, the precise scope and content of the constitutional principle of openness has not been established. Klaaren does provide reference (in footnote 22) to an article in which he subjects the Court’s decision in Independent Newspapers to detailed and careful analysis, and provides deeper discussion of the concept of open justice.9 In this article, Klaaren notes the significance of open justice in South African society, but also that ‘openness has not been given much attention thus far in the South African constitutional project’.10 Nonetheless, all four judgments in the Independent Newspapers case are, in Klaaren’s view, essentially coherent within the framework of open justice.11 Klaaren also sees great significance in the fact that the subject matter of Independent Newspapers were court records and documents in respect of which intelligence agencies had made national security claims – such matters are said to be the ‘heartland’ of open justice, since ‘courts and the public are the classical articulation of civil society’.12

But is it the fact that it is courts or other such bodies acting which is important, or the nature of their actions? Noting that openness was at the core of many (then) current political disputes, Klaaren comments that a relevant aspect of open justice is ‘the extent to which the concept can be used to interrogate the judiciary’s own practice’.13 This statement is echoed in the key paragraph of Klaaren’s original piece, which states that ‘[a] practice of judges related to demands for openness, judicious transparency is a particular form of transparency’ – so it is a ‘judicial practice’ – which is not concerned with the ‘individual dignity of any particular judge’ – so it is not about judicial transparency (as noted above) - but that the ‘manner of implementation’ is ‘what makes it particularly judicious’ – implying that, in fact, it is the nature of the practice, or implementation of the principle of openness, that is what makes it ‘judicious’. In other words, we are back to the plain language meaning: judicious transparency is where a judge applies the principle of transparency in a prudent and sensible fashion, showing ‘good judgment’, that renders it ‘judicious’. It is the juxtaposition of the notion of ‘judicious’ with judicial practice that is confusing. While the argument is framed in

10 Klaaren (n 11 above) 32.
11 As above.
12 Klaaren (n 11 above) 34.
13 Klaaren (n 11 above) 32.
the context of judges being judicious in their application of the principle of transparency, it turns out that since it is the manner of the application and not the identity or office of the implementer of the principle (in Klaaren’s case, the judge), perhaps it could be argued that the concept, if meritorious, could be applied to any context where the principle of transparency, or open justice, was being considered?

This may only serve to take one back to first principles of the right of access to information: that the right is no more absolute than any other right, and that legally permissible exemptions to the right will exist, in the public interest, based on a relationship between the exemption and an understanding of the harm to the public interest that would be caused by disclosure of the information. In this interpretation, it is hard to see what the notion of ‘judicious transparency’ would add. However, Klaaren positions his argument not in terms of the right of access to information, but in terms of the looser idea of ‘transparency’, which may or may not overlap with a legal right of access to information.

Indeed, while it may at first sight appear merely to add to the sense of bewilderment, the final words of the key paragraph in Klaaren’s paper, identified above, may turn out to be the most incisive: ‘As for its substantive value, the substance of that value is indicated by the three rights from which the concept of open justice has been constructed’ (freedom of expression, access to information and access to court). Thus, the idea emerges not so much as a principle or doctrine, but as a value that should infuse judicial decision making when applying the principle of open justice. In other words, stripped down, judicious transparency is the sensible and prudent application of open justice.

2 ‘Judicious transparency’ in practice

Klaaren’s first port of call is the media statement issued by the Constitutional Court announcing its decision to lay a complaint against the Judge President of the Western Cape High Court, Hlophe JP. He notes that Hlophe JP ‘was not appraised of the allegations or their source’; ‘was not asked for his version or comments’; ‘received no effective prior notice of the intention to lodge the complaint’; and ‘was not told of the intention to issues a media statement’. Klaaren notes further that the general public were not given any details about the complaint (the media release having been automatically sent to subscribers to the court’s electronic information system), and were ‘left with nothing more than the knowledge that a complaint with serious implications had been lodged’.

Thus, while the decision of the Court to release the media statement was probably neither unlawful nor improper, it may well have been imprudent and, therefore, injudicious. Indeed, if so, it would have been interesting for Klaaren to apply his own understanding of ‘judicious
transparency’ to the decision. Instead, he turns to two court decisions and, while there are comprehensive and clear explanations as to what the courts found, the analysis only hints at a more interesting engagement and critique. Concerning the courts’ treatment of the issuing of the media statement, Klaaren notes that the judgment of the South Gauteng High Court was ‘somewhat convoluted’, and notes the division of the judges, but does not discuss how either the majority or the minority arrived at their decisions. Beyond noting that the Court ultimately decided in favour of Hlophe JP, no basis is set for the subsequent discussion of the Supreme Court of Appeal (SCA) judgment.

The SCA judgment receives more attention, with Klaaren pointing out that the SCA identified two distinct questions for consideration, one being whether the media statement was unlawful for publishing the fact that a complaint had been made to the JSC and, second, whether Hlophe JP should have been heard prior to the complaint being made. Klaaren observes that the High Court had viewed the laying of the complaint and the issuing of the media statement as one issue, and had thus not dealt with the ‘precise question of publication’. Klaaren comments that the SCA ‘clearly linked’ its decision on the lawfulness of the publication with the issue of the lawfulness of the laying of the complaint, and points out that the SCA’s reasoning is ‘significantly circumscribed’, being focused on the need for public disclosure of the complaint and the circumstances surrounding it, in order to protect the independence of the Constitutional Court and the integrity of the administration of justice. Klaaren notes that ‘[t]his is an argument based on the Court’s core judicial function’, but sees the statement as also suggesting that the Constitutional Court thought it likely that the incidents would become known, and ‘thus did not appear to have total confidence in the informational integrity of its own decision-making processes’. Klaaren concludes his discussion by noting that the appeal turned on the SCA’s finding that the possible truth of the allegation had to be considered in determining the lawfulness of the publication of the media statement, and thus ‘without actually examining or determining the truth of the media statement, the SCA was able to determine the statement not to be unlawful’.

This is an interesting framing of the decision, but unfortunately Klaaren does not develop his analysis of it much further. There are hints at a more provocative engagement with the Court’s reasoning than the paper actually provides. Despite his careful analysis of the decision, Klaaren does not disclose initially whether he agrees with the SCA or not, although in part three of the paper he asserts that ‘[t]he final legal determinations’ – one of which must be the SCA’s decision – ‘were correct as a matter of law’. An analysis of the SCA decision might have benefited from some linking of the decision to previous case law. For example, the lawfulness of the publication of the complaint could have been considered in light of existing South African case law on defamation, but no such context is given. This becomes problematic when Klaaren asserts that the decisions were legally
correct, without having provided the legal basis by which to assess that correctness, nor a clear analysis and explanation of why this is so. A second problem is that a clear link is not drawn between the analysis of the court decisions and the transparency thesis which Klaaren advances.

There is even less critical analysis regarding the decisions relating to the openness of the JSC hearing, with Klaaren contenting himself with setting out a concise overview of the various decisions. But again, more critical engagement with some of the issues would have been welcome. For instance, regarding the court application by various media organisations to open the JSC hearings, Klaaren does not link this to his comments about ‘media transparency’. He notes the JSC’s arguments that the dignity of the process needed to be protected, but does not assess this argument in light of the different theories of transparency which he later discusses. Klaaren then notes Willis J’s finding that the dignity of the judiciary would be enhanced by the hearings taking place in public. What are the implications of this finding for Klaaren’s theories of transparency? Is this an example of ‘media transparency’, since this was the outcome of a review brought by media organisations, and thus presumably according with their vision of transparency? Or is it an example of ‘judicious transparency’? If the latter, why? What distinguishes transparency of either kind in this instance? Similarly, the decision by Malan J\textsuperscript{14} to open a preliminary hearing which the JSC had closed does not receive comment. And yet, one might ask whether this was it a good example of transparency – and of what kind? Were there any special features of a preliminary hearing which might impact on the analysis?

These issues suggest that there are significant unanswered questions about the exact content and parameters of the principle of openness advocated by Klaaren which, regrettably, the paper never delineates amply.

3 Concluding comments

Klaaren disavows any comment on whether or not the laying of the complaint by the Constitutional Court judges was ‘politically wise or ethically necessary’, since the issue is said to be beyond the focus of the paper. Nonetheless, a long paragraph is devoted to it and, in light of this, it was perhaps necessary for the author to give some sense of his own view on whether or not the complaint was advisable or justifiable. Klaaren suggests that the SCA judges indicated that, as an ethical (rather than a legal) matter, the Constitutional Court judges’ actions might be judged in a more negative light. How is this position reconciled with the claim that there is judicious transparency in the issuing of the media statement?

Furthermore, the characterisation of the issue as ‘ethical’ as opposed to legal might itself be open to question. In footnote 31, Klaaren sets out the extract from the SCA judgment in *Langa v Hlophe*, which provides the basis for the suggestion that the complaint might be judged negatively from an ethical point of view. It is notable that the SCA relied on the constitutional injunction that all organs of state, including the judiciary, must ‘assist and protect the independence, impartiality, dignity, accessibility and effectiveness of the judiciary’. This is derived from section 165(4) of the Constitution, and calls into question drawing a distinction between how the complaint is viewed ethically and legally for, if the basis for an argument that the judges’ conduct was ethically questionable was found in the Constitution, is it not the case that such conduct is therefore also legally suspect?

Klaaren concludes that judicious transparency can be seen in the issuing of the Constitutional Court media statement for three reasons: the importance of the subject matter, because it was a ‘statement by judges about a judge and about judging’; the fact that it was done by judges and thus ‘constituted a judicial practice’, rendering the judges potentially liable for defamation; and in the ‘proper context’, beyond the narrow reasoning of the SCA, taking into account the role of national intelligence agencies and constituting ‘a broader attack on judicial independence’. Regarding the latter ground, Klaaren remarks that ‘[t]his is judicious, appropriately not as a judicial function, but rather as a non-judicial decision taken by those entrusted to deciding judicially’. It is unclear how, particularly in light of the arguments he has been advancing up to this point, Klaaren regards a non-judicial decision as being an example of judicious transparency. It may suggest, as certain other parts of the paper do, that what determines judicious transparency is not the nature of the how of transparency is exercised that makes it judicious so much as the identity of who is making the transparency decision.

In which case, while perhaps less intellectually stimulating an exercise, an approach that sought to define the parameters for judicial transparency – such as openness around interviewing and application for judicial office, conflicts of interest and private financial holdings – might have made for a more useful contribution to both the theoretical understanding and application of the principle of transparency at a time when the demand for information disclosure as of right is growing and the dangers of secrecy – in the judiciary as in all sites of public and private power – more and more obvious.

What might the foundations of such a theory of ‘judicial transparency’ be? It is beyond the scope of this paper to develop a detailed theoretical framework, but some basic components may be identified here (in the hope that they might provoke further debate and thought). Firstly, the criteria for judicial appointment ought to be clearly established and openly known. Despite the JSC having been established under the 1996 Constitution,
there remains little transparency as to the criteria it uses for judicial selection.\textsuperscript{15} There is arguably a particular need for transparency in respect of the need to transform the judiciary, which is expressly mandated, in terms of race and gender, in the Constitution. Transparency of the criteria used for judicial appointments allows for informed debate as to the sufficiency of the criteria used, allows those wishing to put themselves forward for nomination to make an accurate appraisal of their prospects for appointment, and facilitates the role of the media in providing information in prompting debate.\textsuperscript{16} Moreover, it is also argued that the quality of decision making is improved when the decision makers are clear on the criteria they are to apply.\textsuperscript{17}

Once appointed, there is a need for transparency in respect of judges’ financial interests, in order to properly regulate and avoid conflicts of interest. Relatedly, there is a need for transparency in respect of the extra-judicial work undertaken by judges, as well as work undertaken by judges after their retirement from the bench. Judicial transparency would also require awareness of judges’ political affiliations – indeed, any factor which might entitle a litigant to request a judge’s recusal ought to be disclosed and available to litigants.

Lastly, a theory of judicial transparency would need to take into account the standards of conduct expected from judges both on and off the bench. And it must do all this whilst fostering the independence of the judiciary, as demanded by the South African Constitution.

Many of these issues have been the subject of considerable attention and often controversy in South Africa, as elsewhere, in recent years. This demonstrates a clear need to establish and define the standards that judges must meet. Unless and until these standards are clearly defined, on the basis of a professional and public consensus that is conspicuous by its absence at present, the underlying causes of many, if not all, the issues Klaaren deals with – not least the divisive controversies surrounding Western Cape Judge-President John Hlophe – will continue to harm the integrity of the judiciary. Injudicious opacity, to coin a phrase, will thus serve to undermine the rule of law.

\textsuperscript{15} S Cowen ‘Judicial selection’ unpublished paper commissioned by the DGRU 7 (Publication forthcoming, 2010).
\textsuperscript{16} As above.
\textsuperscript{17} As above.