## CONTENTS

### THE RULE OF LAW, THE SEPARATION OF POWERS & ANTI-CORRUPTION EFFORTS

- **Pushing the Boundaries: Judicial Review of Legislative Procedures in South Africa**  
  Stephen Gardbaum, University of California  
  Page 1

- **A Tactical Separation of Powers Doctrine**  
  Aziz Z Huq, University of Chicago  
  Page 19

- **Precautionary Constitutionalism, Representative Democracy and Political Corruption**  
  Firoz Cachalia, University of the Witwatersrand  
  Page 45

### CRIME AND PUNISHMENT

- **Plea Bargaining in South Africa: An Economic Perspective**  
  Richard Adelstein, Wesleyan University  
  Page 81

- **Reconciliation as the Aim of a Criminal Trial: Ubuntu's Implications for Sentencing**  
  Thaddeus Metz, University of Johannesburg  
  Page 113

### LAND REFORM, HOUSING & THE SPATIAL EMBEDDEDNESS OF Apartheid

- **Unpacking Section 25: What, If Any, are the Legal Barriers to Transformative Land Reform?**  
  Jackie Dugard, University of the Witwatersrand  
  Page 135

- **From Housing to City: On the Possibilities of the Right to the City in South Africa and India**  
  Ajey Sangai, O.P. Jindal Global University  
  Page 161

- **Dismantling Apartheid Geography: The Limits of the Law**  
  Ralph F Madlalate, Harvard University  
  Page 195

### THE RELATIONSHIP BETWEEN STRUCTURE & CONTENT IN CONSTITUTIONAL ANALYSIS

- **A Service Conception of the Constitution: Authority, Justification, and the Rule of Law in Proportionality Analysis**  
  Richard Stacey, University of Toronto  
  Page 219

- **Forcing the Court’s Remedial Hand: Non-Compliance as a Catalyst for Remedial Innovation**  
  Helen Taylor, University of Oxford  
  Page 247

- **Invalid Court Orders**  
  Mitchell Nold de Beer, Cape Bar  
  Page 283

### AB, REPRODUCTIVE RIGHTS & CHILDREN’S RIGHTS

- **Surrogacy, Geneticism and Equality: The Case of AB v Minister of Social Development**  
  Denise Meyerson, Macquarie University  
  Page 317

- **The Constitution as an Instrument of Prejudice: A Critique of AB v Minister of Social Development**  
  Donrich Thaldar, University of KwaZulu-Natal  
  Page 343

- **The Best Interests of the Child and the Constitutional Court**  
  Meda Couzens, University of KwaZulu-Natal  
  Page 363

### THE LIMITED INFLUENCE OF AFRICAN MORAL THEORY & LAW ON THE CONSTITUTION

- **Kindred Strangers: Why has the Constitutional Court of South Africa Never Cited the African Court on Human and Peoples’ Rights?**  
  Tom Gerald Daly, University of Melbourne  
  Page 387

- **Human Dignity in the Common Law of Contract: Making Sense of the Barkhuizen, Bredenkamp and Botha Trilogy**  
  Hanri Du Plessis, University of South Africa  
  Page 409

### THINGS WE DON’T LIKE TO TALK ABOUT

- **The More Things Change: Multiracialism in Contemporary South Africa**  
  Martin Brassey, University of the Witwatersrand  
  Page 443

- **The Silent Right: Environmental Rights in the Constitutional Court of South Africa**  
  Ruth Krüger, University of the Witwatersrand  
  Page 473

- **Judicial Independence and the Office of the Chief Justice**  
  C H Powell, University of Cape Town  
  Page 497

- **The Rights of the State, and the State of Rights in State Information Technology Agency v Gijima Holdings (Pty) Limited**  
  Robert Hlati Freeman, University of Oxford  
  Page 521

- **Breaking the Silence: The Treatment of Ius Cogens in Zimbabwe Torture Docket and Al Bashir**  
  Mia Swart, HSRC  
  Page 537
Pushing the Boundaries:
Judicial Review of Legislative Procedures in South Africa

STEPHEN GARDBAUM

ABSTRACT: In recent years, the South African Constitutional Court has dramatically shed the general reluctance it had shared with most courts around the world to review legislative processes as distinct from outcomes. In a series of graduated steps culminating in two 2017 cases, the Court has engaged in increasingly robust oversight of various types of legislative procedures. These processes embrace not only the law-making process itself, but also internal National Assembly rules, especially those relating to the National Assembly’s other central function in a parliamentary democracy of holding the executive politically accountable. The article begins with a brief discussion of the background norm of non-intervention in legislative procedures from which the Court has progressively and so notably departed. It then charts the three steps by which this departure has come about, showing how each of them marks a new stage in the degree of judicial supervision. The heart of the article explores whether the Court was justified in taking these steps or was guilty of overreaching. It argues that, although a certain general tension between the separation of powers and rule of law underlies the background norm of judicial non-intervention, in the specific contexts in which these cases were decided, these two values increasingly came together. Indeed, far from violating separation of powers, the Court promoted it when overly concentrated legislative-executive power threatened impunity. Systemic weaknesses of executive political accountability that arises not merely from the existence but the abuse of dominant party status called for novel remedies of the type employed by the Court.

KEYWORDS: legislative process, executive accountability, political process theory, parliamentary government

AUTHOR: MacArthur Foundation Professor of International Justice and Human Rights, University of California, Los Angeles, School of Law.
Email: gardbaum@law.ucla.edu

ACKNOWLEDGEMENTS: Earlier versions of this article were presented at the Constitutional Court Review IX Conference, August 2–3, 2018 and the 2019 ICON-S conference in Santiago. I would like to thank fellow conference participants, three anonymous peer reviewers, and especially Kate O’Regan and Stu Woolman, for extremely helpful comments, suggestions and questions.
Over the past dozen years or so, the South African Constitutional Court has dramatically shed the general reluctance it shared with most courts around the world to review legislative processes as distinct from outcomes. In a series of graduated steps, culminating in the two 2017 cases of United Democratic Movement v Speaker of the National Assembly¹ and Economic Freedom Fighters and Others v Speaker of the National Assembly and Another (‘Economic Freedom Fighters II’),² the Court has engaged in increasingly robust oversight of various types of legislative procedures. These include not only the law-making process itself, but also internal National Assembly rules and mechanisms for the conduct of its business, especially those pertaining to its other key function in a parliamentary democracy of holding the executive politically accountable. Moreover, it has not only found existing promulgated National Assembly rules unconstitutional, as in violation of the constitutional rights of individual members of parliament, but also mandated the creation of others where they do not already exist. Although this series of judicial rulings is an intrinsic part of, and cannot be fully understood apart from, the Court’s broader role alongside other institutions and actors in the ‘politics of accountability’ that ultimately resulted in Jacob Zuma’s resignation and disgrace,³ it is also sufficiently unusual within the narrower frame of separation-of-powers jurisprudence to merit more single-focused or isolated study.⁴

In this article, I first briefly discuss and illustrate the background norm of non-intervention in legislative procedures from which the Court has progressively and so notably departed in recent years. Then, in Part II, I chart the three steps by which this departure has come about, showing how each of them marks a new stage in the degree of judicial supervision. Finally, in Part III, I explore why the Court has been able to take these steps and what the normative justification for them might look like, in terms of such core values of constitutional democracy as the separation of powers and rule of law. Although there is a certain general tension between these two, which underlies and grounds the background norm of judicial non-intervention, I will suggest that in the specific contexts in which these cases were decided, they increasingly came together. Special separation of powers and rule of law problems, arising not merely from the existence but the systematic abuse of dominant party status by the leadership of the ANC, called for special remedies of the type employed by the Court. More broadly, the article contributes to the growing scholarly focus on the general theory and practice of judicial review of legislative processes by analysing its current comparative outer boundary, in South Africa, and proposing a defence of this practice that suggests the need for a friendly amendment to, or extension of, Ely-style political process theory.⁵

2  [2017] ZACC 47, 2017 (2) SA 571 (CC).
4  This view is, I think, supported by the fact that in his outstanding article for the same Rule of Law symposium in this volume, Firoz Cachalia also analyses and seeks to develop an explanatory and normative framework for these and other related judicial rulings. F Cachalia ‘From Aspiration to Realism: Constitutionalism, Judicial Review and the Democratic Process in South Africa’ (2019) 9 Constitutional Court Review – (forthcoming).
5  See text between footnotes 88 and 93 below.
I THE GENERAL NORM OF JUDICIAL NON-INTERVENTION

The starting point from which the recent jurisprudence has progressively departed is the long-standing principle of judicial non-intervention in the internal affairs and procedures of the legislature that is a general characteristic of courts in constitutional democracies. In the common law world, this principle was originally captured in the doctrine of parliamentary privilege, especially as instantiated in Article 9 of the US’s 1689 Bill of Rights, which, inter alia, rendered ‘proceedings of Parliament’ immune from impeachment or questioning in any court of law. Although, of course, this doctrine emerged within a constitutional system gradually evolving towards its central characteristic of the ‘sovereignty of Parliament’ and consequent absence of judicial review of legislation, the two are not necessarily connected. Even within the common law world, the core of parliamentary privilege survived the rejection of parliamentary sovereignty, as evidenced by the adaptation of the language of Article 9 in the United States Constitution exactly one hundred years later and, more recently, in ss 58 and 71 of the South African Constitution. Moreover, the distinction between judicial review of legislative outputs and procedures – of what a legislature has the legal power to do and how it goes about doing it – is reasonably clear in both theory and practice. The substance of parliamentary privilege and the principle of judicial non-intervention it embodies are not, of course, limited to the common law. Most modern democratic constitutions contain similar provisions, providing immunity from court proceedings to members of the legislature for votes cast or statements made in Parliament.

The principle of judicial non-intervention in the internal affairs of the legislature is itself ultimately an implication of the broader, and similarly near-universal, constitutional principle of the separation of powers. This principle requires that ‘each arm of the State has its own domain, in which it is (to varying degrees) a master of its own process.’ The autonomy of legislative proceedings from court oversight is the flip side of the independence of the judiciary from undue or inappropriate legislative interference. Just as it would presumptively violate separation of powers norms for legislatures to prescribe how courts conduct their judicial function, so it would also be for courts to intervene in how legislatures conduct their internal (non-constitutionally specified) proceedings. The separation of powers presumption of institutional autonomy applies equally to inter-branch relationships other than that between the legislature and the judiciary. Parliamentary privilege was also designed to protect legislatures from executive interference, in the form of instigating or threatening prosecutions; and life

---

6 ‘That the freedom of speech and debates or the proceedings of Parliament ought not to be impeached or questioned in any court or place out of Parliament.’ Bill of Rights, 1689, Article 9.

7 US Constitution, Article I, §6 (‘...and for any Speech or Debate in either House, they [Senators and Representatives] shall not be questioned in any other Place.’)

8 ‘Cabinet members, Deputy Ministers and members of the National Assembly (a) have freedom of speech in the Assembly and its committees, subject to its rules and orders; and (b) are not liable to civil or criminal proceedings, arrest, imprisonment or damages for (i) anything they have said in, produced before or submitted to the Assembly or any of its committees, or (ii) anything revealed as a result of anything that they have said in, produced before or submitted to the Assembly or any of its committees.’ Constitution of the Republic of South Africa, 1996 (’Constitution’) s 58(1).

9 See, for example, the German Basic Law Article 46; Spanish Constitution Article 71.

tenure was created in order to protect the king’s judges from the king’s wrath. In addition, many other doctrines, rules, and principles are aimed at constraining one branch of government from overreach into the affairs of another.

As an example of the general reluctance of constitutional courts to intervene in legislative processes and proceedings, even those that engage in robust review of legislative outputs, consider the United States Supreme Court. As is well-known, the Supreme Court and commentators are largely in agreement that judicial review of congressional procedures is more problematic than review of its substantive outcomes, so that almost never is the former undertaken. Rather than direct review of legislative processes as an independent basis for unconstitutionality, the closest courts in the United States typically come is to make them a factor in substantive review. Thus, in determining whether or not Congress is acting within its enumerated authority to regulate interstate commerce, the Supreme Court has held that the existence, nature and quality of congressional findings of a connection between the regulated activity and interstate commerce is a relevant, although not a conclusive, factor. Similarly, where a congressional statute creates a novel legislative procedure for a given subject matter, courts may determine its substantive constitutionality by reference to its consistency with the procedure contained in Article I. Even the few academics who have urged the courts to abandon their resistance to review of process have largely limited themselves specifically to review of law-making rules and procedures. Other types of congressional proceedings are mostly off the table. For example, internal rules of the US Senate, which include such arcane parliamentary manoeuvres as the budget reconciliation procedure that permitted the 2017 Republican tax law to be enacted without the possibility of a Democratic filibuster, as well of course as the filibuster itself, are generally deemed to be non-justiciable. A relevant instance of this position relates to impeachment proceedings brought against two sitting lower federal court judges in the late 1980s. The judges filed suits challenging as unconstitutional the relevant Senate rules under which they had been tried. Although in each case the presiding

---

11 The Act of Settlement 1701 granted royal judges life tenure as distinct from serving at the king’s pleasure, as most previously did.
13 Siman-Tov refers to this as ‘semiprocedural judicial review. Bar-Siman-Tov (note 12 above) at 1924–1925.
17 Originally entitled The Tax Cuts and Jobs Act, the final title was changed as a result of a Senate rule to an Act to provide for reconciliation pursuant to titles II and V of the concurrent resolution on the budget for fiscal year 2018.
18 The filibuster is, of course, (selectively) employed in the Senate’s ‘advise and consent’ function regarding presidential nominations for offices, and not only in law-making debates. On non-justiciability, see A Winkler ‘Is the Filibuster Unconstitutional?’ The New Republic (March 7, 2013); M Miller ‘The Justiciability of Legislative Rules and the “Political” Political Question Doctrine’ (1990) 78 California Law Review 9341.
judge indicated some sympathy to the claims on the merits, both suits were dismissed as non-justiciable. On appeal, the Supreme Court affirmed, holding that impeachment proceedings and procedures were exclusively a political question not entrusted to the courts.

Even the twin pillars of ‘political process’ review in the United States, _Carolene Products_ Footnote 22 and John Hart Ely’s _Democracy and Distrust_,23 still focus on judicial review of legislative outputs, as distinct from processes per se; on what legislatures have done rather than how they have done it. This now seemingly outlier theory famously limits robust or heightened judicial scrutiny to a subset of enacted laws of specific content: those that restrict the democratic process (by, for example, denying some part of the electoral public a vote or voice/free speech) or harm certain racial, ethnic, or religious minorities out of hostility or prejudice, thereby denying equal protection and participation in a representative system.25 Despite its label, the theory does not prescribe judicial review of legislation on procedural grounds alone, much less of non-legislative acts and procedures of the legislature.26

In South Africa, the common law doctrine of parliamentary privilege has, of course, been subsumed under the Constitution and its overarching principle of judicially enforced constitutional supremacy.27 Nonetheless, the core of the doctrine has been constitutionalised in ss 58 and 71,28 in provisions not dissimilar to those in many other modern constitutions. More importantly, the broader principle of judicial non-intervention in internal parliamentary affairs still has force as part of the more general constitutional principle of separation of powers. In _Doctors for Life International v Speaker of the National Assembly and Others_ (‘_Doctors for Life’),29 considered below, the Court was at pains to stress that judicial interference in parliamentary proceedings is inconsistent with the separation of powers doctrine unless mandated by the Constitution.30

In comparative terms, there are a few limited exceptions to the general reluctance of constitutional courts to intervene in legislative proceedings; though the number may be growing worldwide. For example, in 2002, the German Constitutional Court struck down a controversial immigration bill on the basis that the vote in the Bundesrat did not satisfy the Basic Law’s procedural requirement that decisions of the chamber require the consent of a majority of the states, even though the President of the Bundesrat had certified there was a majority (by one state) for the bill. The Basic Law specifies that each state’s votes must be cast as a block vote,31

20 Ibid. See also Miller (note 18 above).
22 _United States v Carolene Products, Co._ 304 US144 (1938).
25 See _Carolene Products_ (note 22 above) at fn 4, sections (2) and (3); Ely (note 23 above) at 103.
26 For more on how my account relates to political process theory, see the Conclusion below.
28 See note 8 above: Section 71 reproduces, for delegates to the National Council of Provinces, the same rights and immunities that members of the National Assembly enjoy under s 58.
30 Ibid at para 37.
31 German Basic Law Article 51(3).
and the Court found that the dissenting vote by one member of the deciding state’s delegation nullified the vote of that state. In 2010, the same court invalidated the Federal Parliament’s recent amendment of the country’s welfare legislation because it deemed that the flawed procedure used to determine the subsistence minimum violated the constitution. In 2017, the Israeli Supreme Court struck down a tax law on procedural grounds alone for violating the right of all legislators to meaningfully participate in the law-making process, when they received the final version of an omnibus tax bill at the last minute. As a final example, the Colombian Constitution expressly grants the Constitutional Court the power to review legislation for errors of procedure as well as substance. In addition to the fairly frequent exercise of this power for legislative infringement of detailed textual procedural rules, the Colombian Court has also invalidated a tax law for violating the unwritten principle of ‘minimum public deliberation.’ This latter invalidation occurred when the government, in an attempt to reduce the budget deficit, added increased VAT rates on certain necessities at the last minute with no notice to legislators just before the final vote. The Colombian Court’s interventions have been premised on an attempt to bolster the quality of deliberation within a largely dysfunctional legislature, in the light of the clear intent of the framers of the 1991 Constitution to try and overcome a long history of overly powerful presidents and rubber stamp legislatures.

II THE THREE-STAGE DEPARTURE FROM THE NORM

In South Africa, the Constitutional Court’s departure from this norm of non-intervention has taken place in three distinct, increasingly robust stages between 2002 and 2017. Starting out with judicial review of legislative Acts for potential violations of constitutionally required procedures at stage one, the Court then moved into the more unorthodox terrain of reviewing internal parliamentary rules for infringements of the constitutional rights of individual Members of Parliament at stage two, and finally, at stage three, began reviewing whether and how (ie, the procedures by which) the legislature performs its critical non-legislative function in a parliamentary democracy of holding the executive politically accountable. This third, most recent, stage is highly unorthodox by comparative standards. As a quintessentially political function, judicial oversight and review would typically be viewed as illegitimate overreach.

Stage one consisted of two major cases reviewing the constitutionality of legislative Acts, the first for what was done and the second for what was not, decided in 2002 and 2006 respectively. In United Democratic Movement v President of the Republic of South Africa, the Constitutional

---

33 The ‘Hartz IV’ Decision, BVerfGE, Judgment of the First Senate of 9 February 2010 – 1BvL 1/09 – at paras 1–220.
34 Kwantiski v The Knesset, HCJ 10042/16 (20017). The decision is discussed in I Bar-Siman-Tov ‘In Wake of Controversial Enactment Process of Trump’s Tax Bill, Israeli SC Offers a Novel Approach to Regulating Omnibus Legislation’ I-CONnect (December 13, 2017).
35 Constitution of Colombia Article 241.
37 Decision C-776 of 2003 (per Justice Manuel José Cepeda Espinosa). For further discussion of the decision, see Espinosa & Landau (note 36 above) at 318–323.
Court reviewed four Acts of Parliament, including two constitutional amendments, permitting limited floor crossing by members of national, provincial, and municipal legislatures that were challenged by the opposition UDM as being in violation of various procedural requirements under the Constitution. With respect to the two constitutional amendments, the claim was that their passage required meeting, but did not satisfy, the higher thresholds of ss 74(1) or (2), as amendments of the founding values of the Constitution in s 1 or the bill of rights in chapter 2, rather than the general or default threshold of a two-thirds vote of the National Assembly under section 74(3). With respect to one of the statutes, the Membership Act, the claim was that the Constitution’s temporary licence to change the initial ban on floor crossing by ordinary statute had expired, so that another constitutional amendment was now required. Indicating that it has ‘little if any’ role in reviewing the substance of constitutional amendments (including ones that amend the so-called basic structure), as distinct from whether amendments are ‘passed in accordance with the prescribed procedures and majorities’, the Court unanimously rejected the claim that the floor-crossing amendments triggered the higher threshold as inconsistent with either s 1’s founding values of multiparty democracy or the rule of law, or the political rights contained in Chapter 2. Rather, these values and rights do not require or mandate an anti-defection provision, as both the temporary licence and many other democratic constitutions testify. The Court unanimously accepted UDM’s temporal argument regarding the Membership Act. Although the product of the Constitution’s unusual, but by no means unique, special entrenchment provisions or ‘tiered constitutionalism’, as a species of procedural review of constitutional amendments, the Court’s role was well within the comparative mainstream. Similarly, its apparent rejection of both substantive review of constitutional amendments and inquiring into the motives or consequences of ANC support for floor-crossing, are also evidence of judicial modesty.

Four years later, in Doctors for Life, the Court took the bolder step of invalidating four health statutes because the National Council of the Provinces had failed to fulfil its constitutional obligation to ‘facilitate public involvement’ as a requirement of the law-making process under ss 72(1)(a) and 118(1)(a). Acknowledging that separation of powers and the autonomy of Parliament mean the judiciary should not interfere in its processes unless mandated to do so by the Constitution, the Court drew the outer boundaries of judicial power at determining whether there has been the degree of public involvement in the law-making process required by the Constitution, but leaving to parliamentary discretion how the duty is fulfilled.

40 This higher threshold is 75 per cent.
41 UDM (note 39 above) at para 12. The Constitutional Court’s jurisdiction to review the validity of constitutional amendments is granted by s 167(4).
43 For an argument that the Court ought to have taken into account the impact of floor crossing on strengthening the ANC’s dominant party status, see S Choudhry “He Had a Mandate”: The South African Constitutional Court and the African National Congress in a Dominant Party Democracy’ (2009) 2 Constitutional Court Review 1.
44 Constitution s 118(1)(a) reads, in relevant part: The National Council of Provinces must – (a) facilitate public involvement in the legislative and other processes of the Council and its committees. Constitution s 72(1)(a) reads, in relevant part: ‘A provincial legislature must – (a) facilitate public involvement in the legislative and other process of the legislature and its committees.’
45 Doctors for Life (note 29 above) at para 67. A similar decision, involving a provincial legislature’s approval of a constitutional amendment, was reached in Matatiele Municipality v President of the Republic [2006] ZACC 12
obligation against the legislature also calls for a second-order reasonableness standard on the ‘whether’ question: has Parliament acted reasonably in discharging its duty to provide a reasonable opportunity for public participation.\(^{46}\) Despite this relatively permissive approach, the majority concluded that holding no public hearings at all had been unreasonable under the circumstances.

The next stage in the progression occurred in 2012–2013, in two cases reviewing and invalidating internal National Assembly rules (rather than legislative Acts) for violating the constitutional rights of opposition party members. In Oriani-Ambrosini, MP v Sisulu, MP Speaker of the National Assembly (‘Oriani-Ambrosini’),\(^{47}\) the Court held that National Assembly rules requiring MPs to secure the permission of the Assembly, through one of is prescribed committees, before they may prepare and introduce legislation, violates their individual right to do so under ss 55(1) (b) and 73(2).\(^{48}\) Although s 57 empowers the National Assembly to create rules covering its legislative business, this authority is textually conditioned on respecting the participation of minority parties ‘in a manner consistent with democracy’.\(^{49}\) It does not permit such rules to negate what the Court interpreted as the constitutional right of individual MPs to initiate and to introduce legislation by imposing what are effectively substantive (rather than simply procedural) limitations on its exercise. The following year, in Mazibuko v Sisulu and Another (‘Mazibuko’),\(^{50}\) the Court applied essentially the same reasoning to what it interpreted as the individual right of an MP to schedule a motion of no-confidence in the President of the Republic under s 102(2).\(^{51}\) National Assembly rules delegating to a committee the power to decide whether to schedule such a motion for debate in the Assembly effectively impose substantive limitations on the right and are inconsistent with the Constitution. Once again, although National Assembly rules may prescribe the process for vindicating a right, they may not thwart or frustrate it: ‘We may not hold that an entitlement that our Constitution grants is available only at the whim or discretion of the majority or minority of members on a committee.’\(^{52}\) By seven votes to four, the Court ordered a suspended declaration of invalidity for six months to remedy the defect. In these two cases, the Court reviewed the constitutionality of internal National Assembly Rules and found the procedures that they establish impermissible fetters on the individual (rather than collective or institutional) rights of MPs – it perhaps generously read the relevant sections to contain.

\(^{46}\) Ibid.


\(^{48}\) Constitution s 73(2) reads, in relevant part: ‘In exercising its legislative power, the National Assembly may – (a) consider, pass, amend or reject any legislation before the Assembly; and (b) initiate or prepare legislation, except money Bills.’ Constitution s 55(1) reads, in relevant part: ‘Only a Cabinet member or a Deputy Minister, or a member or committee of the National Assembly, may introduce a Bill in the Assembly...’

\(^{49}\) Constitution s 57: reads, in relevant part: (1) The National Assembly may – (a) determine and control its internal arrangements, proceedings and procedures; and (b) make rules and orders concerning its business, with due regard to representative and participatory democracy, accountability, transparency and public involvement. (2) The rules and orders of the National Assembly must provide for – ...(b) the participation in the proceedings of the Assembly and its committees of minority parties represented in the Assembly, in a manner consistent with democracy.

\(^{50}\) [2013] ZACC 28, 2013 (6) SA 249

\(^{51}\) Constitution s 102(2) reads: ‘If the National Assembly, by a vote supported by a majority of its members, passes a motion of no confidence in the President, the President and the other members of the Cabinet and any Deputy Ministers must resign.’

\(^{52}\) Mazibuko (note 50 above) at para 58.
The third and most recent stage (2016–2017) involves a trio of cases in which judicial review of parliamentary processes went beyond both legislative Acts and internal procedural rules for consistency with constitutional rights. Here, the Court intervened in the distinctive legislative function in a parliamentary democracy of holding the executive politically accountable. Ordinarily, political accountability stands in contrast with legal accountability, and so more-or-less by definition excludes judicial oversight and review.

In *Economic Freedom Fighters v Speaker of the National Assembly* (‘Economic Freedom Fighters I’), the Court unanimously held that the National Assembly has not only the political power and function, but also a constitutional obligation, to hold the President accountable. Further, it had violated this duty by not facilitating and ensuring compliance with the legally binding recommendations of the Public Protector involving Jacob Zuma’s private home – Nkandla. Instead, the National Assembly (‘NA’) had passed a resolution absolving the President based on a report by the Minister of Police, a member of the President’s cabinet. It thereby effectively flouted its constitutional obligation, as well as the rule of law. The Court derived this obligation from s 55(2), which directs the NA to create mechanisms ‘to ensure that all executive organs of the state are accountable to it.’ As in *Doctors for Life*, the Court drew the boundary of judicial power at the ‘whether’, rather than the ‘how’, question:

> It falls outside the parameters of judicial authority to prescribe to the National Assembly how to scrutinize executive action, what mechanisms to establish...for the purpose of holding the executive accountable and fulfilling its oversight role...Ours is a much broader and less intrusive role. And that is to determine whether what the National Assembly did does in substance and in reality amount to fulfilment of its constitutional obligations. That is the sum-total of the constitutionally permissible judicial enquiry to be embarked upon. And these are some of the ‘vital limits on judicial authority and the Constitution’s design to leave certain matters to other branches of government.’

The Court concluded unanimously that the NA’s constitutional obligation had been violated.

In *United Democratic Movement v Speaker of the National Assembly* (‘UDM’), decided in June 2017, the Court set aside the Speaker’s ruling that she had no power to call for a secret ballot on a motion of no-confidence in the President. In response to the Speaker’s view that neither the Constitution nor the rules of the NA provide for secret ballot, the UDM Court held that as the Constitution is silent on the voting procedures, a secret ballot is permissible and the NA has, through its Rules, effectively delegated to the Speaker the decision as to what procedure to use, including a secret ballot. In addition, the Court held that the Speaker’s decision must be supported by a proper and rational basis and made to facilitate the effectiveness of parliamentary accountability mechanisms. Recall that in *Doctors for Life*, the Court stated that it could only intervene in parliamentary proceedings when the Constitution mandated it

---

54 The report containing the recommendations was entitled *Secure in Comfort: Public Protector’s Report on Nkandla: Report by the Public Protector on an Investigation into Allegations of Impropriety relating to the Installation and Implementation of Security Measures by the Department of Public Works at and in respect of the Private Residence of President Jacob Zuma at Nkandla in the Kwa-Zulu Province Report No: 25 of 2013/4 (19 March 2014).*
55 The legally binding nature of the Public Protector’s recommendations was the first, and perhaps even more important, part of the Court’s ruling in *EFF I*. Woolman (note 3 above).
56 Ibid at para 93 (quoting *Doctors for Life* (note 29 above) at para 37).
58 *EFF I* (note 53 above) at paras 86–88.
to do so. Here such a mandate is not as easy to find. By acknowledging constitutional *silence* on
the issue of voting procedures and effectively intervening to interpret NA rules to the Speaker,
the Court seems to be extending its role. Moreover, reviewing internal procedural decisions of
the chief parliamentary officer under the general principle of legality59 is arguably in significant
tension with parliamentary autonomy. It is true that the Court refuses UDM’s request to
affirmatively order the Speaker to hold a secret ballot, as ‘no legal power exists for such a radical
and separation of powers-insensitive move’;60 but under the circumstances, such an order was
probably unnecessary. As the Court concluded: ‘[N]ow that it has been explained that she
[the Speaker] has the power to do that which she is not averse to, she has the properly-guided
latitude to prescribe what she considers to be the appropriate voting procedure under the
circumstances.’61 The circumstances, of course, included this unanimous judicial ‘guidance’.

Finally, in *Economic Freedom Fighters II*,62 the Court took its intervention into the
procedures surrounding Parliament’s task of holding the executive accountable to a new and
higher level. Although the opposition parties had moved to impeach the President under
section 89(1),63 and the motion was debated and defeated on a secret ballot ordered by the
Speaker in the light of UDM, the Court held that the NA had nonetheless failed to fulfill its
implicit constitutional obligation under the section to make rules creating a specially tailored
process for impeachment. The more general ad hoc committee procedure used by the NA for
the motion was, according to the majority, constitutionally insufficient. The procedure was
also found inadequate in a second way: rather than the required antecedent judgment that one
of the grounds for impeachment exists, such as a ‘serious violation of the Constitution or the
law’, the NA impermissibly held a one-step debate on the motion to impeach.64

The vigorous dissent by the Chief Justice and three colleagues helpfully highlights how the
*Economic Freedom Fighters II* majority went beyond the judgments in the previous two cases,
which he had written for a unanimous court. According to the dissent, whereas in *Economic
Freedom Fighters* and *UDM*, no specific rule or procedure was prescribed by the Court, here
there is. But the Court should only be asking whether a satisfactory procedure for impeachment
was in place, and not the best one. As the Chief Justice wrote:

This time around, we are even specific about size, representations, procedure, provision for the
entirety of the process, avoiding abuse of majority representation, institutional predetermination of
grounds before debating and voting on impeachment. That, in my view, is an unprecedented and
unconstitutional encroachment into the operational space of Parliament by judges. . . . There exists
no jurisdiction in the whole world, that I am aware of, where a court has decided for Parliament
how to conduct its impeachment process. Respect for separation of powers explains why this is so.65

---

59 The principle, as developed by the Constitutional Court, requires that exercises of public power comply with
minimum standards of lawfulness, reasonableness and fairness. See, for example, C Hoexter ‘The Principle of
60 *EFF I* (note 53 above) at para 92.
61 Ibid.
62 *EFF II* (note 2 above).
63 Constitution s 89(1) reads, in relevant part: ‘The National Assembly, by a resolution adopted with a supporting
vote of at least two-thirds of its members, may remove the President only on the grounds of (a) a serious violation
of the Constitution or the law; (b) serious misconduct, or (c) inability to perform the functions of office.’
64 *EFF II* (note 2 above) at para 180.
65 Ibid at para 254.
In prescribing ‘what mechanisms to establish’, the majority also went beyond, or perhaps showed the impossibility of maintaining, the Chief Justice’s distinction in *Economic Freedom Fighters I* (going back to *Doctors for Life*) between impermissible judicial review of how Parliament holds the executive accountable and obligatory judicial review of whether it fulfils its constitutional obligation to do so.

In sum, by the end of 2017, judicial review of parliamentary procedures included not only the various components of the law-making process, National Assembly rules for violations of constitutional rights, decisions of the Speaker made under or applying these Rules, and whether Parliament has fulfilled its constitutional obligations as interpreted by the Court, but also what mechanisms and processes these obligations require. In terms of institutional independence and the separation of powers-inspired limits on judicial power, one would be hard pressed to state precisely what remains of the autonomy of Parliament vis-à-vis the courts apart, somewhat ironically, from the outcomes of judicially-approved procedures.

### III ACCOUNTING FOR THE DEPARTURE

In this section, I first want to address the explanatory question of why the Court has been able to intervene in parliamentary processes in this increasingly robust way. Then, the remainder of the section will focus on the normative issue: what might a justification for the Court’s intervention look like in terms of separation of powers and rule of law values? Are these two necessarily in tension, so that advancing the latter has come at the expense of the former? Is there a way to understand the Court’s intervention as promoting, rather than (perhaps justifiably) trenching on, separation of powers?

Both sides of the ‘constraint’ identified by Theunis Roux as characteristic of the Court – South Africa’s general legal culture and the political environment in which it operates – have created the space for, or facilitated, the recent jurisprudence of legislative process. Working within a relatively formalistic or legalistic culture that tends to treat as illegitimate judicial recourse to broader moral and substantive norms, the text of the Constitution provides unusually rich resources in the relevant sections to support the Court’s work, in that it contains a comparatively large amount of detail and specificity on legislative functions, powers and, especially, duties. This, of course, is not to say that the judicial rulings were compelled by the text, or are even the best among competing reasonable interpretations, but rather that the text can plausibly be read to support them. In almost every case, the leading opinion was able to cite (and interpret) a relevant section of the text rather than rely exclusively on more general structural norms of democracy, parliamentary government, minority rights, separation of powers, etc.

---


67 In *Doctors for Life*, the Court relied on s 72(1)(a) and public participation. In *Oriani-Ambrosini*, its decision hinged upon ss 55(1)(b) and 73(2) regarding the rights of MPs. In *Mazibuko*, it found support in s 102(2) on motions of no confidence. In *EFF I*, it banked on s 55(2)’s requirement that the legislature must hold the executive accountable. In *EFF II*, it found succour in s 89’s provisions on removing the President. Indeed, the Court has held that the values contained in Section 1 of the Constitution, which include democracy and the rule of law, although fundamental and giving substance to all the provisions of the Constitution, do not give rise to discrete and enforceable rights in themselves. *Minister of Home Affairs v NICRO* [2004] ZACC 10 at para 21.
way, the Court could reasonably claim – and be seen – to be doing ‘lawyers’ work’,\textsuperscript{68} despite the increasingly obvious political consequences.

At the same time, the political constraints on the Court have loosened over the course of this time period, coinciding with the playing out of the Nkandla and other recent corruption scandals,\textsuperscript{69} as the dominance of the ANC has seemingly declined. This is evidenced by the sweeping losses in the 2016 municipal elections, the forced resignation of Jacob Zuma, and at least talk that the ANC could conceivably have faced defeat at the national elections in 2019. This larger story, of the ‘politics of accountability’ and the ‘catalysing effect’ of the roles of media, civil society groups, opposition parties, Public Protector and judicial decisions in defending the rule of law and ‘resetting’ democracy, has been masterfully told by Stu Woolman in his contribution to these pages last year.\textsuperscript{70} For current purposes, it seems reasonably clear that the trajectory of the Court decisions discussed in the previous section were both in part cause and effect of the erosion of ANC dominance. In 2002, or even 2012, the Court did not have the political space to act as it did in 2017.

If these two factors help to explain the Court’s increasingly robust interventions in parliamentary processes, why it was able to play this role, what if anything might justify it? In acknowledging the tension between separation of powers and the rule of law in this context, is the Court best seen as perhaps justifiably jettisoning the former in order to protect the latter in the face of increasing brazen displays of impunity? Although this is not of course how the Court did or could present its case for intervention, I believe that separation of powers and the rule of law in fact come together, rather than diverge, to provide a justification for the Court’s role.

To make this case, it is necessary to be reminded that there are two relevant sides of the institutional separation of powers, and not only one. As briefly discussed in part I above, the first side, and the one most obviously at stake in judicial consideration of legislative processes, is the mutual independence and autonomy of these two branches of government from each other. Just as the legislature is presumptively prohibited from intervening in judicial processes and functions as potentially undermining that part of the separation of powers usually referred to as the independence of the judiciary, so too the courts are constrained to respect the autonomy of Parliament, as an independent – and, in a parliamentary system, the only directly elected – branch of government. The role of constitutional courts in enforcing the supreme law limits on (and, where they exist, the constitutional obligations of) the legislature is expected to take this principle into account, hence the general tension between these aspects of the separation of powers and rule of law.

\textsuperscript{68} ’Lawyer’s work’ is a reference to US Justice Scalia’s critique of the popular legitimacy of the Supreme Court’s work product when its decisions make little or no reference to the constitutional text. A Scalia ’Originalism: The Lesser Evil’ (1989) 57 University of Cincinnatti Law Review 849, 854.

\textsuperscript{69} For more on the ‘state capture’ scandal involving the Gupta family’s influence over Jacob Zuma to steer procurement decisions and political appointments for their benefit, as well as the practices of other major private firms designed to advance their interests and their profits, see the Public Protector’s second major report on Zuma-era corruption: Office of the Public Protector, South Africa State of Capture: Report on an Investigation into Alleged Improper and Unethical Conduct by the President and Other State Functionaries relating to Alleged Improper Relationships and Involvement of the Gupta Family in the Removal and Appointment of Ministers and Directors of State Owned Enterprises resulting in Improper and Possibly Corrupt Award of State Contracts and Benefits to the Gupta Family Businesses, Report No. 6 of 2016/17 (October 14, 2016), available at saflii.org/images/329756472-State of Capture.pdf.

\textsuperscript{70} Woolman (note 3 above).
But the second side of the separation of powers involves legislative-executive, rather than legislative-judicial, relationships. In presidential democracies, this aspect of separation of powers is institutionalised in separated powers, with politically independent legislative and executive branches typically each having a role or veto in certain shared functions, such as law-making. Although the original Madisonian, friction-based model of ‘ambition being made to counteract ambition’ may have had its rough edges both smoothed down and further sharpened in regular cycles by the modern cross-branch partisanship of political party and party discipline, the relative independence of each resulting from the logic of separate popular elections survives. Thus, whether the re-election prospects of legislators are enhanced or diminished by actively supporting a president of their own party (or opposing one of the other) is generally determined by the contingencies of politics at any given time, not structure. By contrast, in parliamentary democracies the system of partially ‘fused’ rather than separated powers resulting in large part from holding a single election for both branches, means that, on the one hand, the executive governs significantly through the legislature but, on the other, the separation of powers is served by means of the executive’s political dependence on and responsibility to parliament. Stated crudely, what the parliamentary executive ‘gains’ in terms of political power compared to the presidential through its typically majoritarian control of the legislature, it also ‘loses’ by being dependent on the legislature for continuance in office. Although, therefore, parliamentary executives and legislatures are not designed or structured to be fully independent of each other, nonetheless the legislature must be sufficiently independent of the executive to be able to play its key separation of powers role of holding the executive politically accountable to it. If it were not, partially fused power would become fully fused, or concentrated, power, the very opposite of separation of powers.

In the modern context, many parliamentary democracies around the world have come to the conclusion that, sometimes, judicial intervention may be required to supplement or bolster the capacity of legislatures to play their allotted role in making the parliamentary version of separation of powers work. The nature, scope and limits of such judicial intervention have varied, depending on a range of factors. In terms of the requisite independence of Parliament and possible judicial supplementation, there are basically four different scenarios.

The first may be called the classical, and is best represented by the British parliamentary system of the nineteenth century, before the development of the modern political party system. Here, political accountability of the executive to a legislature consisting of relatively independent Members of Parliament – elected to office at least as much for their personal abilities, connections and records as their party labels and able to deliberate collectively about the public interest – was a robust feature of parliamentary government. As long as it continued to enjoy the confidence of a majority of such members, the government exercised significant legislative powers in addition to executive ones (reflected in the concept and prioritising of ‘government bills’), meaning that there was a partial fusion of the two functions and not only personnel. But the possibility of losing

73 As a result, parliament tends to be a ‘policy-influencing’, rather than a ‘policy-making’ legislature (as in a system of presidential government). R Craig ‘Could the Government Advise the Queen to Refuse Royal Assent to a Backbench Bill?’ UK Constitutional Law Blog (22 January 2019).
that confidence was a real one, so that executive dependence on – and political accountability to – Parliament reflected the practice and not merely the theory of governance. As evidence, between 1832 and 1945 there were seventeen successful votes of no confidence in the British Parliament resulting in the resignation of the prime minister. Since 1945, there has only been one.\textsuperscript{75} Indeed, not since 1895 has a majority-party government been voted out of office by Parliament. In this classical period, the traditional separation of power between King and Parliament\textsuperscript{76} was reconstituted by the new separation between the King’s government in Parliament and the other, non-ministerial members of the legislature, a separation which, although it involved some overlapping office holding, was still institutional in nature and designed to structure executive-legislative relationships in a way that maintained some significant independence of Parliament. Moreover, this second side of the separation of powers not only did not require, but would have condemned, any intervention on the part of the courts. Political accountability was not only sufficient to ensure this aspect of the separation of powers, but, in contrast to legal accountability, it also reflected and expressed the distinctive role of Parliament as the historic protector of liberty. This was the heyday of ‘the political constitution’.\textsuperscript{77}

The second scenario is the contemporary competitive party parliamentary democracy, which displaced the classical following the rise and triumph of the modern political party system. Compared to the first, it is characterised by more limited separation of powers and political accountability to parliament. In versions with two-party (and so, usually, first-past-the post voting\textsuperscript{78}) systems, the typical electoral outcome of a single majority party in Parliament without majority electoral support has reversed the dependency of the executive on the legislature. As party political affiliation, rather than individual record and qualities, became all-important for voting purposes, the re-election prospects of government and individual Members of Parliament were increasingly intertwined at the one and only, dual-function election. Thereafter, the primary task of ‘backbench’ majority party legislators was no longer to check and hold the government politically accountable but to maintain it in office, increasingly at all costs. This new electoral interdependence of government and Parliament resulting from the modern party system has really become the dependence of the latter on the former. The executive dominates Parliament through the whip system in order to maintain necessary party unity in the face of continuous partisan attack by the minority second, or opposition, party: the alternative government-in-waiting. Accordingly, without any significant alteration of formal institutional powers or relations, the entire working of the parliamentary system in Britain changed – from a more divided or pluralistic to (usually) a strongly unified governance structure – as a result of the rise of the modern party political system in the first half of the twentieth century. And this arrangement was duplicated in other Westminster-style systems. In his classic work on political parties, Maurice Duverger described the result of this change as follows:

\begin{quote}
Officially Great Britain has a parliamentary system...In practice the existence of a majority governing party transforms this constitutional pattern from top to bottom. The party holds in its hands the essential prerogatives of the legislature and the executive...Parliament and Government
\end{quote}

\textsuperscript{75} Prime Minister James Callaghan resigned in 1979.

\textsuperscript{76} A Tomkins \textit{Public Law} (2003).


\textsuperscript{78} Following Duverger’s rule (or law): (1) one-seat districts with plurality/majoritarian electoral rule tend to reduce the number of parties to two; and (2) multi-seat districts with proportional representation tend to be associated with more than two parties. M Duverger \textit{Les Partis Politiques} (1951); M Duverger \textit{Political Parties: Their Organization and Activity in the Modern State} (1954).
are like two machines driven by the same motor – the party. The regime is not so very different in this respect from the single party system. Executive and legislature, Government and Parliament are constitutional facades: in reality the party alone exercises power.\textsuperscript{79}

Within multiparty versions, usually resulting from some form of proportional representation electoral system,\textsuperscript{80} there is typically somewhat more ‘divided’ government, not because of institutional separation but because of the balance of party politics. Two or more parties share executive power in a coalition government, and the executive as a whole is somewhat more dependent on the legislature as it often cannot take continued support more or less for granted as is usual in two-party systems. Smaller parties in multiparty systems tend not to be ‘all-purpose’ entities whose main political interests lie in supporting or opposing the government whatever its position, but are more issue driven and so more selective in their voting – whether inside or outside the coalition. The religious parties in Israel are a classic example of this phenomenon. Accordingly, overall there is less concentration of power and a less comprehensive fusion of executive and legislative functions.

Nonetheless, in both two-party and multiparty competitive democracies, the relative loss of parliamentary independence in the face of party government and growing ‘executive supremacy’ has meant that the capacity of legislatures to subject the executive to meaningful political accountability has been in secular decline everywhere. Political accountability is no longer sufficient for the oversight and checking function that the second side of the separation of powers originally assigned to the legislature. As a result, in this new context, political accountability of the executive for its actions has been supplemented by legal accountability. This, I believe, is a large part of the reason for the rise of judicial review, not only of administrative acts, but also of executive-driven legislation that the parliamentary world has seen since 1945 and again after 1989.\textsuperscript{81} But note than such judicial review, like the political review it supplements, is focused on outputs rather than processes. It serves as a partial substitute for the continuous political accountability of the executive for its legislative and administrative actions in between the longer electoral cycles of direct accountability to voters.

The third scenario is the dominant party parliamentary democracy. The major difference from the ordinary competitive party system of the second scenario is that it is not merely Parliament’s capacity to hold the executive accountable that is at risk, but the existence of any meaningful form of political accountability where the possibility of electoral turnover in power is small or effectively zero. In a competitive party system, in addition to grounding periodic accountability to the electorate, such rotation in office means that while executive dominance may be a constant, it is exercised by alternating actors and parties over time, so that there is at least temporal separation or diffusion of power – even if with a lag – and no complete and indefinite concentration or capture of state power by one group (or subgroup).\textsuperscript{82} By contrast, in a dominant party democracy, such temporal concentration significantly exacerbates the separation of power problem by posing a threat to the independence not only of Parliament

\textsuperscript{79} Ibid at 124.
\textsuperscript{80} Ibid.
\textsuperscript{81} S Gardbaum ‘Separation of Powers and the Growth of Judicial Review in Established Democracies (Or Why has the Model of Legislative Supremacy Mostly Been Withdrawn From Sale)’ (2014) 62 American Journal of Comparative Law 613.
\textsuperscript{82} In some unstable competitive party democracies, the same political actors often remain in office despite frequent changes of government. Nonetheless, these are coalition governments, lacking the concentrated power of a dominant party government.
but of all institutions, including the courts, and thereby beyond political to any form of accountability and the rule of law. Where a court has the political space to do so, this threat may justify additional scope for judicial intervention, including attempts to bolster the effectiveness of the legislature’s accountability mechanisms, as compared with the second scenario.83

And yet, despite the risks that inhere in the type, they may not be realised. Moreover, in a constitutional democracy, if one party earns the long-term trust and support of the majority of voters through free and fair elections, this legitimises its dominance despite the risks and, other things being equal, should constrain the types of judicial interventions in the name of the separation of powers that are deemed proper. Accordingly, where a dominant party essentially or mostly plays by the rules (even ones it largely wrote) and does not succumb to the pathologies posed by the form,84 at least generally, then the scope for judicial intervention in an attempt to defend, boost or maintain the independence of Parliament in particular (although other institutions as well), while perhaps greater than in a competitive system, should be circumscribed.

The final scenario may be termed systematic abuse of dominant position. Here, the risks posed by the form are in fact realised generally. Not only is the potential for concentration of power over time achieved, but its typical pathologies appear more regularly and systematically: corruption, impunity, lack of accountability. Once abuse takes the form of illegal acts and the ruling group’s dominance is employed to evade responsibility for them, then separation of powers and the rule of law join forces to justify greater judicial intervention in an attempt to open up the blockages to both political and legal accountability. In this scenario, the autonomy of a perennially captive Parliament becomes a weapon used by the party leadership to resist, not to promote, its being held to account, and judicial intervention in legislative processes is geared towards loosening party control and boosting the independence of opposition voices to help secure its key separation of powers function. In other words, in this extreme dysfunctional scenario, the larger separation of powers goal of resisting the fact and consequences of concentrated power over time may justify incursions of the principle’s more normal institutionalisation in a democracy. Judicial intervention to bolster Parliament’s ability to play its central role in holding the executive accountable, especially for violating the law, may be called for. This judicial oversight of legislative procedures as an attempt to remedy the problem of institutional dysfunction is similar to what has transpired in Colombia, as mentioned in Part I above, even though the relevant function and context – though not the underlying concern with separation of powers – differ. In Colombia’s presidential system, the problem is weakly institutionalised political parties, rather than a dominant one,85 and the dysfunction relates to the deliberativeness of the law-making process rather than holding the executive accountable.

It is this final scenario that has been playing out in South Africa over these past few years, as clear and incontrovertible evidence of a shift from (1) the mere existence of a dominant

83 Of course, despite such potentially greater justification, typically courts have less political space to intervene under this scenario, so that the risks of doing so, in terms of threats to judicial independence, are also likely to be greater. See Roux (note 66 above). For Asian examples, see P Yap Courts and Democracies in Asia (2017). See also S Gardbaum ‘Are Strong Constitutional Courts Always a Good Thing for New Democracies?’ (2015) 53 Columbia Journal of Transnational Law 285.
84 For a concise summary of these pathologies, as well as early evidence of occasional or partial realisation, see Choudhry (note 43 above).
85 Espinosa & Landau (note 36 above).
political position, to (2) occasional or partial abuse of that position, and finally (3) to systematic and widespread abuse by the party leadership has emerged. The argument is not intended to provide a specific justification for, say, the majority versus the dissenting judgment in Economic Freedom Fighters II, but rather a more general or functional justification of the entire trend of the Court’s jurisprudence as a whole as it has increased the scope and intensity of its review of legislative processes of all types. Both sides of the separation of powers are aimed at preventing a common problem, the undue concentration of political power and its consequences, and when there is a major risk of systemic failure, available resources can be diverted in a crisis to the immediate source of danger. This is how judicial intervention in parliamentary processes to bolster the legislature’s ability to hold an abusive executive accountable, especially for violations of the law, should be seen. Special separation of powers problems call for special remedies.

IV CONCLUSION

Within a system of constitutional supremacy, there is a tendency for the rule of law to become equated with judicial supremacy or the rule of judges. No person is above the law, but the Constitution is above the ordinary law, and its meaning and application are commonly perceived to be the ‘peculiar province of the courts’. The principle of the separation of powers, which guards against the consolidation of power by any one person or institution, protects against this tendency by imposing limits on judicial power that typically take a number of forms. One of these boundaries is the institutional autonomy of the legislative branch of government from judicial oversight of its internal proceedings, if not necessarily (or any longer) its external outputs. Accordingly, a certain tension between the rule of law and the separation of powers is a normal feature of modern constitutional democracies, especially where (as in South Africa) the supreme law not only grants powers to, but imposes affirmative obligations on, the legislature. Should these obligations be understood, interpreted and enforced more through the lens of the rule of law or the separation of legislative and judicial power?

This normal tension, however, may largely disappear in circumstances where the threat of consolidated power and its consequences comes from a dominant political party in long-term control of the executive and legislative branches, as well as other state institutions designed to disperse and check power, and hold it accountable. Within parliamentary systems, separation of powers is importantly manifested and operationalised through the political accountability and responsibility of the executive to the legislature. Where abuse of dominant position results instead in the weaponisation of the legislature to help create impunity and resist accountability – especially for criminal violations by party leaders – the rule of law and separation of powers come together to justify judicial intervention to bolster the ability of a parliamentary legislature to fulfil its distinctive function. The same would be true in reverse. Where abuse takes the form of systematically stripping the courts of their independence and packing them with party loyalists to the point of constitutional or criminal impunity, both rule of law and separation of powers may render certain unorthodox legislative intervention to resist and counter such moves legitimate. In both cases, the real issue will often be less the legitimacy than the practical likelihood of such measures in a context of abusive dominance. But where

86 For more evidence on this subject, see the two Public Protector reports on Nkandla and state capture (notes 54 and 69 above).
sufficient institutional independence has survived to make them possible, they are justified and proportionate as a means to the goal of saving constitutional democracy.

In the final sentence of *Democracy and Distrust*, John Hart Ely wrote that ‘constitutional law appropriately exists for those situations where representative government cannot be trusted, not those where we know it can’. 88 Systematic abuse of dominant position is one of these ‘situations’. 89 More generally in this article, I have offered an ‘elaboration’ 90 and extension of the political process or representation-reinforcement theory of judicial review that Ely constructed on the foundation of Footnote 4. This is so in three ways. First, despite its ‘political process’ label, the existing theory still essentially addresses judicial review of legislative outcomes, *what* was done rather than *how*, albeit focusing on different types of laws (those denying a vote or voice, or harming ‘discrete and insular minorities’ 91 out of prejudice or hostility) than those of the rival ‘substantive values’ approach. By contrast, the account presented in this article seeks to provide a justification of judicial review of legislative processes per se, both statutory and non-legislative acts and procedures of the legislature. In other words, political process review and legislative process review are not identical. Second, to the extent it does focus on actual procedures, Ely’s theory looks primarily to the ‘external’ processes and modes of democratic participation by the people, especially in terms of voting and voice in free and fair elections; whereas my account looks to the ‘internal’ processes of, and modes of participation in, the elected legislature. Finally, the article identifies some of the distinctive problems and potential ‘malfunctioning [of] the political market’ 92 within a parliamentary constitutional democracy as compared with the presidential one that Footnote 4 and Ely presumed.

Nonetheless, although falling within none of the Footnote 4 categories or Ely’s articulation and extended defence of them, and so neither a restatement nor a necessary implication of this theory, the account provided here undoubtedly falls within its general spirit. For, like Ely, it is centrally concerned with providing means of resistance to prevent ‘the ins . . . choking off the channels of political change to ensure that they will stay in and the outs will stay out.’ 93

---

88 Ely (note 23 above) at 183.
89 I am not arguing that systematic abuse of dominant position is the *only* scenario that potentially justifies judicial intervention in legislative procedures of the type undertaken by the South African Court. Even without establishing political dominance, a governing party may engage in forms of ‘abusive constitutionalism’ that similarly threaten separation of powers, rule of law, and constitutional democracy: as arguably has occurred in Poland since 2015. That said, the persistence of competitive party democracy – and thereby political accountability – will likely mean that the judicial role ought to be more circumscribed (as generally under the second scenario) and this type of intervention retained as a last resort.
90 Ely (note 23 above) at 181: ‘The elaboration of a representation-reinforcing theory of judicial review could go many ways, and Chapters 5 and 6 are obviously just one version.’
91 *Carolene Products* (note 22 above) at 152, fn 4.
92 Ely (note 23 above) at 103.
93 Ibid.
A Tactical Separation of Powers Doctrine

AZIZ Z HUQ

ABSTRACT: This essay explores the possibility that courts can play a role in arresting damage to constitutional democracy and in hindering processes of democratic backsliding. To that end, it examines closely four decisions in which the Constitutional Court has responded to state capture as a threat to ‘constitutional democracy’. Each case reflects an effort either to stymie President Zuma’s determination to entrench his hold on high office or to buttress the attempts by opposition parties to oust him from power. This coterie of cases are rightly perceived as component parts of an ongoing accountability process in which the Court employs tactical strategies that rely upon the use of structural constitutional commitments rather than clearly enunciated principles or provisions in the basic law.

KEYWORDS: comparative constitutional law; separation of powers; democratic decline

AUTHOR: Frank and Bernice J Greenberg Professor of Law, Mark Claster Mamolen Teaching Scholar, University of Chicago School of Law. Email: huq@uchicago.edu

ACKNOWLEDGEMENTS: Thanks to Stu Woolman, Yasmine Dawood, Theunis Roux, and two anonymous reviewers for helpful criticism.
I THE PROBLEM OF STATE CAPTURE IN SOUTH AFRICA

In Democratic Alliance v Speaker of the National Assembly and Others, the Constitutional Court of South Africa began its analysis of the free-speech rights of parliamentarians by declaring simply that ‘South Africa is a constitutional democracy.’ Madlanga J did not define the term. Nevertheless, his opinion took it as an axiomatic basis of South African constitutional adjudication. The following year, Mogoeng CJ began his opinion in United Democratic Movement v Speaker of the National Assembly and Others by using very similar language. This time, however, the Court added some words of definition. As relevant here, its guiding gloss echoed Abraham Lincoln’s famous Gettysburg Address: A ‘constitutional democracy,’ wrote Mogoeng CJ, is a ‘government of the people, by the people and for the people through the instrumentality of the Constitution.’ This is rather good as rousing rhetoric. It is perhaps less satisfying as a measure for legal interpretation. After all, it is not at all clear how the definition picks out a discrete class of cases in which constitutional democracy is imperiled. So long as there is a government in power with some sort of claim to democratic legitimacy derived in some fashion from the 1996 Constitution, it might be argued that ‘constitutional democracy’ of some kind obtains. It is also not clear how this touchstone of ‘constitutional democracy’ could usefully illuminate constitutional adjudication. In cases concerning the allocation of powers and privileges between different constitutional institutions, a domain sometimes known as the ‘separation of powers,’ for example, the definition provides essentially no real help in understanding ambiguous constitutional provisions without more analytic work. Yet Mogoeng CJ and Madlanga J would not have conjured a concern with constitutional democracy had not that value been imperiled in some sense. They also would not have offered it as a lodestar in their opinions had they not believed that the term in some meaningful way could explain and guide their legal analysis. I believe they were correct on both counts. As a result, their opinions invite an inquiry into how a constitutional court can defend or protect constitutional democracy – even if the bare invocation of principle provides little by way of an answer.

The nature of the then-relevant threat to the integrity of constitutional democracy in South Africa is widely known. A perceptive commentator, writing roughly contemporaneously to the United Democratic Movement judgment, summarized the South African political situation at the time in the following terms:

[Jacob] Zuma’s presidency has been disastrous. Corruption has run riot, led by the example of the President’s own family and their friends and benefactors, the Gupta family, an immigrant Indian family whose tentacles are everywhere in government departments and state-owned industries…. In effect, the state has been criminalized. Officials who refuse to allocate contracts in the right direction will receive death threats, and as there were more than two dozen political murders last year, no one believes that this is only bluffing. Confidence and trust in government have never been lower, not even under apartheid.

The origin of the crisis that has thrown ‘constitutional democracy’ into such parlous conditions, Samuel Issacharoff has argued in these pages, is political rather than legal or constitutional in genealogy. In recent years in South Africa, Issacharoff contended, ‘the dominant party [the African National Congress, or ANC, became] the center for all political

1 [2016] ZACC 8, 2016 (3) SA 487 (CC) at para 11.
2 United Democratic Movement v Speaker of the National Assembly and Others [2017] ZACC 21, 2017 (5) SA 300 (CC).
3 Ibid at para 1.
and economic dealings with the government. The effects of a dominant party take several forms. The most important was a profusion of graft and systematic corruption. During the tenure of Thabo Mbeki, an ‘ANC party state’ developed in which party loyalists were assigned to high posts in public office, parastatals came under party control rather than state control, and ANC elites increasingly dominated the ‘commanding heights’ of the private economy. In the Zuma presidency that followed, the merger of the state with private interests accelerated. The state was captured by a group of businesses owned and operated by three brothers sharing the Gupta name, who manipulated state business to their personal advantage. This often involved the steering of public contracts to preferred businesses owned by the Guptas at exorbitant rates in exchange for kickbacks to ministers and other officials making allocative decisions. Ministers who declined to cooperate with the Guptas were quickly relieved of their duties and office. As a result of ineffectual or corrupt presidential leadership, a raft of structural, macroeconomic problems accumulated. State-owned enterprises such as Eskom, Denel, and the South African Broadcasting Corporation, for example, fell into dysfunction and courted bankruptcy. As RW Johnson has argued, the compromising of large state enterprises imperilled the state’s access to international credit markets. Eskom’s debt contracts include cross-default clauses. Under these clauses, Eskom’s bankruptcy automatically results in all loans to South African State-owned enterprises (SOEs) being called in. Looting one SOE is thus linked by law to the prospect of a generalised financial crisis that the ANC is not well positioned to parry.

Such endemic corruption and graft preclude the delivery of some basic government services. The macroeconomic strains generated by these inefficiencies are likely to generate even sharper cuts in public services in the future. Relatedly, a generalized failure of law-and-order has led to a sharp spike in crime and vigilantism. This violence often had a political cast. Deadly violence against critics of the ANC, as well as civil-society figures who had the temerity to complain of corruption, spread rapidly. Surveying the resulting lawlessness and impunity, some commentators have proposed a new concept of ‘violent democracy’ to capture South Africa’s predicament. The ordinary processes of political feedback, which might curb some of these

---

7 ‘South Africa’s Public Protector Finds “State Capture” by the President’s Pals’ The Economist (5 November 2016); L Chutel & L Kuo ‘What the “State Capture” Report Tells Us about Zuma, the Guptas, and Corruption in South Africa’ QZ Africa (3 November 2016).
11 Ibid.
trends, have too often been partly stymied as a result of party loyalty or corruption. The ANC rank-and-file, for example, has at some pivotal moments protected corrupt leaders such as Zuma from censure or punishment, using complaints about ‘white monopoly capitalism’ to draw attention away from their malfeasance.14

The net result of these trends, according to a pivotal report by Thuli Madonsela, the former Public Protector of South Africa,15 was a phase shift from diffuse, pervasive graft into a more systematic phenomenon of ‘state capture’, in which the state effectively operated as a vehicle for a narrow, unrepresentative set of private interests acting without democratic authorization.16 While the evidence of state capture documented by Madonsela was particularly vivid, South Africa is hardly alone in facing this cluster of problems. Even before Zuma came to power, political scientists had identified a relationship in African democracies between the concentration of political power (particularly in the office of a president) and pathological forms of clientelism.17

At first blush, the problem of state capture exemplified by the Zuma presidency might seem intractable. It is not at all clear how the tools of constitutional law and institutional design might remedy it at all for three reasons. First, the problem of state capture does not arise simply because one office (the presidency) wielded too much control over fiscal and procurement decisions (although that may make it worse).18 Rather, as depicted by Madonsela’s ‘State of Capture’ report, and as theorized in Issacharoff’s work, state capture arises in part because of the brute electoral dominance of one party. The latter, suggests Issacharoff, is a necessary if not sufficient condition for state capture. In the case of the ANC, that dominance reflects the pivotal role that the party played in the independence movement, as well as the axial role of its former leader, Nelson Mandela, in the transition from Apartheid. In the absence of the ANC, the 1996 South African Constitution would not exist in its current form. However powerful the party’s claim to democratic legitimacy in the aftermath of independence, the very magnitude of its political dominance has opened a pathway toward state capture.19

Second, one analysis of endemic corruption suggests that the latter arises in the context of accelerated capitalist development, in which government officials inevitably have the power to

14 Bracking (note 8 above) at 170–171.
15 President Zuma initially applied to block the publication of the report, apparently in the hope that it would be revised or edited by Madonsela’s successor. He withdrew his application, however, and in November 2016, the Pretoria High Court ordered the report’s release. See L Wolf ‘The Remedial Action of the “State of Capture” Report in Perspective’ (2017) 20 Potchefstroomse Elektroniese Regsblad 1, 4.
18 For an argument to that effect, see M Mkhabela, ‘Zuma May Be Gone, But How Do We Avoid a Repeat of the Disaster?’ News24 (6 September 2018). RW Johnson also blames the specific leadership failures of various ANC presidents. Johnson (note 10 above). The growth of presidential authority may also impede development of a public culture of democratic expectations and participation. YL Morse ‘Presidential Power and Democratization by Elections in Africa’ (2018) 25 Democratization 709.
19 For an insightful, and early, argument that the ANC would present a threat to the persistence of constitutional democracy by dint of its dominance, see H Giliomee, J Myburgh, & L Schlemmer ‘Dominant Party Rule, Opposition Parties and Minorities in South Africa’ (2001) 40 Democratization 161, 162.
marshal and channel resources, and hence defalcate state funds. But if widespread corruption is associated with a particular phase of economic development, it may not be avoidable through legal means; and ceasing to develop economically is simply not on the cards.

Third, state capture involves a pervasive failure of integrity on the part of high-level officials. This is not so much an institutional design problem as a problem of culture. If a sufficiently anomic attitude suffuses the whole ANC party structure (which has been in power at the national level and in many subnational jurisdictions since 1994), it seems quite unlikely that any institutional fix will prevent state capture. If a sufficient proportion of officials wielding state power are indifferent to the risk to constitutional democracy, if they become sufficiently inured to the public’s legitimate demands for effective state functioning for all, and if they are willing to engage in self-serving conduct without regard to the cost to the nation’s economic and political stability, then it is hard to see how any rule of constitutional law or institutional device could make a difference. This is not a novel point: theorists from Max Weber onward have recognized that a measure of official acceptance and internalization of legal rules is necessary for the law to have effective force. More recently, theorists of constitutional law have derived from the English legal philosopher H.L.A. Hart the powerful hypothesis that the legitimacy and stability of a constitutional system depends on the attitudes and investments of some combination of elites and ordinary citizens. Absent these orientations toward legality and law, no legal system gets off the ground or stays in motion.

But if law rests on the sociological fact of internalization, particularly by state officials, and upon a positive attitude toward legality, it is circular to call upon law as a solution when that internalization, and the corresponding official attitudes, fail. Rather, a Weberian (or Hartian) lens suggests that it may be wiser to cultivate the appropriate sociological conditions in which internalization occurs than to chisel precisely the corners and edges of government institutions. Consider in this regard, the effort of the American constitution’s drafters to elicit certain ‘institutional’ dispositions on the part of national office holders, as a means of deracinating them from the fiery passions of factional passion. The Madisonian project of generating institutional loyalties as a predicate for restraining intragovernmental competition, however, is today considered a failure by most constitutional theorists. There is no contemporary successor to this Madisonian theory that promises to reinfuse government with integrity and legality through clever constitutional design. Constitutions, we now think, cannot make people good.

So there are ample reasons for doubting the efficacy of law or constitutionalism as a complete remedy for state capture. Yet this perhaps moves too fast. The conclusion that laws

---

22 For a recent recapitulation of this argument, see R Fallon Law and Legitimacy in the Supreme Court (2018). For Hart’s canonical description of law as resting upon the ‘customary rule of judges and other officials’, see HLA Hart The Concept of Law (2012), 317.
23 But the idea of cultivating certain dispositions to promote constitutionalism is not entirely alien to constitutional law. Consider, for example, Philip Bobbitt’s idea of an ‘ethical’ form of constitutional argument, which ‘relies on a characterization of [national] institutions and the role within them of the [national] people.’ P Bobbitt Constitutional Fate: Theory of the Constitution (1982), 95. For a South African take on this proposition, see S Woolman ‘On the Reciprocal Relationship between the Rule of Law and Civil Society’ (2015) Acta Juridica 374.
24 For a reconsideration of this project, largely associated with James Madison, as well as suggestions as to how it might be partially redeemed, see D Fontana & A Huq ‘Institutional Loyalties in Constitutional Law’ (2018) 85 University of Chicago Law Review 1.
and institutions are alone insufficient to prevent state capture does not also mean that laws and institutions cannot ever make a difference. Perhaps they cannot guarantee absolutely that state capture, or some other threat to constitutional democracy, will not arise. Law, after all, cannot always prevent the emergence of a dominant party. Nor can it wholly control the forces unleashed by a sudden exposure to global markets.

Nevertheless, law and constitutions might play a role in managing the resulting threats to democracy. Savvy institutional design might change the rate at which state capture takes hold, allowing countervailing political forces time to marshal. It might also mitigate the risk of state capture’s worst-case scenarios. Indeed, the possibility of successful judicial interventions in respect to social pathologies of a different sort was posited by Theunis Roux in 2003. Roux’s core claim was that the Court had quickly become ‘adept’ in the role of ‘legitimatior of the post-apartheid social transformation project’. In 2009, Roux identified the Court’s separation-of-powers jurisprudence as an instrument to ‘minimize the impact’ of controversial cases on the tribunal’s ‘institutional security.’ Roux’s aim was to illustrate how separation-of-powers ideas could be deployed strategically to facilitate the Court’s role in the constitutional architecture. The function played by that doctrine in the decisions under review here is rather different, insofar as it relates to the stability of democracy as a going concern. Nevertheless, my analysis is consistent with that earlier claim, and might profitably be read as an extension of Roux’s pathmarking work.

Another domain in which the role of courts in protecting constitutional democracy is the recent literature on ‘democratic backsliding’. This literature crisscrosses law, political science, and sociology. It aims to understand, and craft effectual responses to, the electoral victories of populist leaders in Poland, Hungary, Venezuela, Bolivia, Russia, Turkey, and the United States (among other countries). Many of these leaders have either attempted to dismantle, or succeeded in dismantling, the likelihood of subsequent democratic rotation. They have neutered checking institutions such as courts and ombudsmen, and silenced dissenting voices in the media and academia. Their legalistic toolkit is novel. In the twentieth century, antidemocratic behaviour generally took the form of coups or abuses of emergency powers. Today, democratic backsliding typically takes place within a framework of laws and constitutional continuity. The process of backsliding, moreover, has typically been incremental, rather than immediate. Decline is a process and not an event. Recent studies suggest that it would be foolish to expect constitutional courts to provide a comprehensive prophylactic against backsliding risk: there are many instances in which elected branches can outflank such tribunals by moving first or by simply stacking the bench.

when courts will succeed or fail in this enterprise. More modestly, I hope to explore one case study where judicial action has contributed, at least temporarily, to democracy’s defence.\(^{30}\)

There are many similarities between the problem of state capture and the prospect of populist degradation of constitutional norms. Studies of the latter process suggest that legal and institutional arrangements can be important, if not always decisive, frictions. Legal rules matter because the processes of institutional decomposition extends over time through the manipulation of legal and constitutional rules; and the details of those rules turn out to be consequential. Constitutional design can create protective redoubts for political factions committed to democracy and a strong form of the rule of law. These institutional shelters enable resistance to democratic backsliding that would otherwise be impossible. Such resistance might not succeed – but without clever constitutional design, it is simply impossible. This dynamic is most clearly illustrated in several cases of democratic ‘near misses’ in countries such as Sri Lanka, Colombia, and Finland. In all those countries, a stable democratic order has come under sudden attack, but has subsequently recovered (at least in the medium-term). In each of these three cases, the Constitution or laws created institutional shelter for resistance to backsliding. Specifically, unelected actors empowered by law to act as election administrators, as judges, or as military commanders took advantage of their structural independence from direct political control to oppose backsliding.\(^{31}\) To be sure, the operation of those institutions could not alone have saved democracy. Rather, their intervention created political space for other, democratic forces to organize successfully, and ultimately prevail in a subsequent election. It is possible that what works to curb populist backsliding also works to constrain the threat to constitutional democracy from state capture.

This essay explores the possibility that courts can play a role in arresting damage to constitutional democracy, much as courts have hindered processes of democratic backsliding in other contexts. To that end, it examines closely four decisions in which the Constitutional Court of South Africa has responded to state capture as a threat to ‘constitutional democracy.’ Each of these cases concerns an effort either to stymie one of President Zuma’s efforts to entrench himself, or else an oppositional effort to oust him from power. One of the four cases was decided after Zuma resigned in February 2018. I include it here because it concerns Zuma’s prosecutorial appointments, which were part of his entrenching strategy. These cases are perceived as component parts of an accountability process. Stu Woolman has argued in these pages that the Constitutional Court was a key player in a larger political dynamic by which Zuma was ‘ignominiously’ forced from the presidency.\(^{32}\) That it was a process rather than an event, notwithstanding the early and ample evidence of Zuma’s improprieties, reflects

---

\(^{30}\) Of course the possibility exists that the kind of malleable, tactical jurisprudential instruments described here can be used for anti-democratic ends. But whereas I am skeptical that more formalist approaches can ever be effective means of democracy preservation, at least tactical approaches can be used for salutary ends.


\(^{32}\) S Woolman ‘A Politics of Accountability: How South Africa’s Judicial Recognition of the Binding Legal Effect of the Public Protector’s Recommendations Had a Catalyzed Effect that Brought Down a President’ (2018) 8 Constitutional Court Review 158, 159. Woolman notes that the Court’s ruling ‘rattled’ the ANC’s hegemony and had ‘discernible knock-on effects’ such as the setbacks of the ANC during 2016 municipal elections and internal ANC political processes that brought Ramaphosa to power. Ibid at 185.
the ANC’s frequent resistance to accountability, and at times its resistance to the Constitutional Court itself.35

I consider these four decisions to analyse how they employed the Constitution against state capture: Did the Court apply a consistent theoretical touchstone in analysing legal questions? Did it consistently favour one branch (or one institution within a branch) over others? Was its aim to elicit solely long-term dynamics? The answer to these questions is no: the Court’s interventions were methodologically varied; they helped different entities; and they tacked between short-term and long-term effects. Indeed, an overarching lesson to be drawn from these cases is that democracy’s defence is a tactical, rather than a strategic, challenge. Rather than articulating a singular theory or jurisprudential decisions, the Court’s decisions are better understood as fluid and contingent responses to specific political risks. Threats to constitutional democracy are fluid and themselves the results of strategic action. Hence, it is not surprising that a court seeking to defend democracy pursuant to the rule of law would be ill-served by a single, fixed theoretical framework. Nor is that goal best served by a singular effort to defang one ‘most dangerous’ branch.34 Indeed, at certain moments the appropriate object of reformist attention is not the state at all, but the dominant party’s internal dynamics. When a single party is sufficiently dominant, its character and trajectory shape the state’s character and trajectory. Moreover, the kind of needful remedies will vary from one occasion to the next. At times, the most needful interventions are local and immediate in effect; at times, they will have little or no direct effect, but will profoundly alter the strategic options for both the defenders and the foes of constitutional democracy. The Constitutional Court’s jurisprudence in this field, in short, illustrates both the possibility and the virtues of a tactical approach to the separation of powers as a safeguard of constitutional democracy. It shows the virtue of setting deep structural principles to one side, and simply doing what works.35

To be clear, I do not claim that these judgments are the only decisions of the Constitutional Court that, in some fashion, have been responsive to the problem of state capture. At other times, the Court has also struck down legislation that diminished the ability to create and to maintain a genuinely independent anti-corruption agency;36 has limited parliamentary delegations to the presidency to extend the Chief Justice’s term of office;37 and has overturned a presidential decision appointing a new National Director of Public Prosecutions.38 The decisions analysed here are simply the rulings most closely tied to the specific threat to constitutional democracy posed by President Zuma. Nor do I claim that the ‘strategic’ quality of the Court’s decisions, which I emphasize here, is somehow unprincipled or unjudicial. To the contrary, I emphasize that methodological pluralism and doctrinal flexibility are directly tied to the higher-level

34 For a genealogy and useful repurposing of this concept, see MS Flaherty ‘The Most Dangerous Branch’ (1995) 105 Yale Law Journal 1725.
36 Glenister v President of the Republic of South Africa & Others [2001] ZACC 6, 2011 (3) SA 347 (CC).
pursuit of democracy-preservation – akin to ‘neutral principles’ – at a doctrinal level. This methodological pluralism and doctrinal flexibility is, of course, neither necessary nor sufficient for the more general structural task of defending a constitution’s core moral commitments.

I begin by offering a parsimonious definition of ‘constitutional democracy’ drawn from my earlier work with Tom Ginsburg on the cognate idea of ‘liberal constitutional democracy’. This definition supplies a more secure point of reference for analysis. It is broadly consistent with the concerns raised by Mogoeng CJ and Madlanga J, and helps explain the relationship between constitutional democracy and corruption. After a brief account of separation-of-powers jurisprudence from the 1990s and early 2000s, I turn to the four recent Constitutional Court decisions touching on state capture and the Zuma presidency, focusing in particular on the way in which the Court’s interventions relate to the project of maintaining constitutional democracy. These decisions illustrate the virtue of a tactical separation of powers. I conclude by stepping back briefly and reflecting on the implications of the (at least temporary) success of the Constitutional Court’s intervention for the ongoing efforts to theorize the separation of powers as a general constitutional concept.

Before offering any further analysis, I feel it important to offer an important caveat: I am not a South African lawyer, and have no particular feel for the fabric of its jurisprudence. Nor (to my dismay and chagrin) have I had the chance to spend sufficient time in South Africa to be able to claim any familiarity with the felt political dynamics and patterns of national political life. The analysis that follows, accordingly, should be understood as reflecting the written products of the Constitutional Court, as well as the available secondary literature, evaluated from a comparative and theoretical perspective. I make no pretence at an evaluation of the nation’s constitutionalism from an internal perspective.

II A WORKING DEFINITION OF CONSTITUTIONAL DEMOCRACY

It is useful to begin by offering a thick account of ‘constitutional democracy.’ By using the term ‘constitutional democracy’, I believe that Mogoeng CJ and Madlanga J had something in mind akin to what Tom Ginsburg and I in a recent book discuss under the rubric of ‘liberal constitutional democracy.’ Mogoeng CJ and Madlanga J are fairly read as implying the broadly ‘liberal’ character of the Constitution, at least in the sense Ginsburg and I use that term. Its omission from their opinions, therefore, is not a strike against this definitional alignment. The idea of liberal constitutional democracy, moreover, illuminates the relationship of state capture to the prospect of democracy backsliding.

In a book and a number of scholarly articles, Ginsburg and I have characterized ‘liberal constitutional democracy’ as comprising three distinct, but interlocking system-level elements. The first of these is a democratic electoral system. Most importantly, this involves periodic free-and-fair elections in which the modal adult can vote, and after which a losing side concedes power to the winning side. The second prong of the definition comprises the particular liberal rights to speech and association that are closely linked to democracy in practice. Without free speech and association, it is hard to see how opposition parties can coordinate and persuade voters to remove

40 This definition draws on various works previously cited (note 27 above). Our definition was parsimonious because we aimed to provide a working basis for analyzing the processes of democratic backsliding, and believed that a thicker description courted the risk of becoming mired in the extended, and perhaps interminable, debates over what should count as a democracy.
incumbents from office. Finally, and most unusually, the definition looks for a level of integrity of law and legal institutions sufficient to allow democratic engagement without fear or coercion. Democracy, we have argued, requires institutional mechanisms to maintain regularity and legality in elections and media regulation. It needs a robust, relatively autonomous and professional, bureaucracy to prevent incumbent self-dealing or entrenchment. This need is particularly acute in respect to institutions that manage the democratic process and the criminal justice system. These are the parts of the state most apt to be used unfairly against electoral competitors. Liberal constitutional democracy is thus a system-level quality that requires all three of these more precise institutional predicates before getting off the ground.

As a system-level quality, it is not easy to precisely quantify. Part of the difficulty of quantification flows from the fact that none of the three institutional predicates that we describe are ever likely to be perfectly achieved. All democracies, however well-functioning, likely fall short in one way or another in respect to these three institutional building blocks. None, however, can work without a measure of all three of the values we describe. And we think it is plausible to make judgments about the existence and direction of liberal constitutional democracy, even if the latter cannot be precisely measured.

Once constitutional democracy is defined in this fashion, the link to state capture intimated by Mogoeng CJ and Madlanga J becomes easier to see. Endemic corruption is causally related to systemic and disabling dysfunctions in the basic democratic mechanisms that link the public to the state.\textsuperscript{41} State capture’s evisceration of the administrative rule of law in turn undermines both liberal rights and the possibility of fair elections by fostering a violent democracy in which the law-enforcement institutions cannot be trusted.

Consider a number of ways in which this can happen. First, state capture undermines the efficacy of channels linking public preferences to political choices. Most generally, a government that is unable to satisfy the basic demands of its citizens for public goods such as safety, macroeconomic stability, and services such as power and water is not a government that can sustain confidence in democracy as a system. Catastrophic economic mismanagement creates conditions for democratic backsliding. Economic crises since the Great Depression have been linked to the rise of anti-democratic populist movements.\textsuperscript{42} On gaining power, populist parties tend to dismantle many of democracy’s necessary institutional supports by removing checks on executive power, hobbling courts and oversight bodies, and capturing the free press and the academy.\textsuperscript{43} Indeed, as discussed above, this dynamic is precisely what has recurred, and become entrenched, in many nations since the recession of 2008–2009.

Both populist governance and state capture can inflict a blizzard of small, but cumulatively crippling, harms to democracy. The result is democratic death by a thousand cuts. A political landscape of endemic corruption is thus incompatible with the motivational basis of democratic participation because it yields an ‘unjustified disempowerment’ of citizens who are not direct


\textsuperscript{42} Barry Eichengreen, \textit{The Populist Temptation: Grievance and Political Reaction in the Modern Era} (2018). If liberal democracy tends to founder under conditions of macroeconomic strain, and economic crisis is sufficiently predictable, then liberal democracy necessarily has an expiration date – even though it is not to blame for macroeconomic crisis. A less pessimistic view is that liberal democracies need to be shored up in anticipation of inevitable economic crisis.

beneficiaries of ongoing graft.\textsuperscript{44} Where people cease to believe that public decisions are based on public-regarding reasons, as opposed to private gain, they become disillusioned with the process of democratic debate and persuasion.\textsuperscript{45} The public sphere is also robbed of its function when its participants come to believe that their arguments no longer have any connection to the actual exercise of state power. In addition, there is a bilateral relationship between state capture and democratic decay. On the one hand, a dominant party engages in state capture as an alternative to competition in the democratic sphere. On the other hand, state capture motivates a dominant party to adopt rhetorical tactics that degrade further the relationship between the public and its elected representation. For example, the ANC has employed rhetorical tropes that demonize whites to sustain public support in the absence of any concrete policy achievements, and against the backdrop of collapsing public services.\textsuperscript{46} In Zimbabwe, such tactics have led to violence and dispossession. While the risk of overt violence against whites as a consequence of this sort of political rhetoric is hard to quantify, the illiberal distortion of public debate makes it more difficult for the polity to focus upon matters of real public importance. Indeed, the whole point of ginning up anger at foreign elites is to distract the public from deep failures of governance. Such ‘democratic pollution’ presents arguably as a great a risk to the possibility of democratic rotation as formal, legal censorship. Just as much as a prior restraint, it impedes the effective operation of public debate and political association.\textsuperscript{47}

In sum, different threats to liberal constitutional democracy target distinct elements of its tripartite institutional foundation. State capture, understood as an acutely endemic species of corruption, undermines directly the administrative rule of law and corrodes the quality of the public sphere. In so doing, it directly and indirectly saps the efficacy of both fair elections and also liberal rights of speech and association. As Mogoeng CJ and Madlanga J suggested, the South African example is an important reminder of a pluralism of democracy’s institutional foundations. Such pluralism reduces the chances of systemic failure, even if it remains possible that a default in one local domain can have concatenating, systemic consequences for democracy as a whole.

III THE CONSTITUTIONAL COURT CONFRONTS STATE CAPTURE: FOUR PERSPECTIVES

This part examines the ways in which the Court interpreted the South African Constitution’s division of powers between branches and other institutions in order to promote constitutional democracy. I focus in particular on the crucial two years prior to Zuma’s resignation in February 2018. I discuss four cases below. Three of them precede Zuma’s fall. The fourth followed that event, but was deeply implicated in Zuma’s fate, and reflects an effort to unravel the pernicious elements of his legacy. How in these cases, I ask, did the Court further the cause of constitutional democracy? What kind of vision of the institutional relations of constitutional

\textsuperscript{44} Ibid 329.
\textsuperscript{46} For examples, see Bracking (note 8 above) at 171.
\textsuperscript{47} Tim Wu has argued that effective political control is achieved more effectively through polluting the public sphere than by formal censorship legible as such. T Wu ‘Is the First Amendment Obsolete?’ in D Pozen (ed) Knight Institute Emerging Threats Series (2017). Wu focuses on the ability of internet content providers to manipulate users’ attention. The form of democratic pollution with which I am concerned here is somewhat different in form and vector.
bodies underwrote its decisions? By examining decisions that were part of a successful defence against democratic backsliding, I hope to offer insight into the role that courts can play in shoring up constitutional democracy.

As a threshold matter, it is worth underscoring that the opinions discussed in this part are consistent with a longer jurisprudential trajectory in which the Constitutional Court has approached the separation of powers in a purposive manner that combines doctrinal and methodological suppleness with an eye to larger, systemic ends. In the First Certification Judgment, for example, the Court refused to endorse a static, rigid, and textualist approach, and instead embraced an approach that emphasized functional principles that could emerge through specific applications and the workings of institutional life. The separation of powers in South African law, it explained, ‘evidence[s] a concern for both the over-concentration of power and the requirement or energetic and effective, yet answerable, executive.’ It was, the Court later elaborated, ‘a delicate balancing, informed both by South Africa’s history and its new dispensation’ between restraining and enabling impulses. The same idea was endorsed again in 2001 and then in 2002. Anticipating some of the themes in the cases discussed below, the Court has also vigorously endorsed the functional independence of Chapter 9 institutions from ‘national executive control’ such that ‘they should be and should manifestly be seen to be outside government.’ In short, there is a precedential foundation for the approach followed in these four cases. What is of interest is therefore not the fact of doctrinal novation, but how this ‘distinctively’ South African approach has been adapted to the question of state capture.

In the four cases examined here, the Constitutional Court has addressed questions about the allocation of institutional power between and within the branches of the national government. These cases have arisen in the context of contestation over how (and for whom) state power is exercised. In particular, they all concern President Zuma’s efforts to remain in office despite allegations of misconduct. The decisions were all steps in a larger process that led to President Zuma’s departure from office in February 2018. Other elements of the process had nothing to do with the courts. For example, Zuma’s March 2017 effort to install ‘a tame Gupta protégé’ as finance minister was taken as a ‘clear sign that even the national Treasury was about to be looted,’ and led to an overnight collapse in the value of the Rand. It brought businesses, trade unions, and the Communist Party into accord that Zuma had to be dispatched. The subsequent publication of the ‘State of Capture’ report also catalysed street protests with ‘huge crowds of trade unionists and workers … chanting anti-Zuma slogans in the street and demanding his resignation.’ Still, informed commentators underscore the important role that the Constitutional Court’s decisions played in the lead-up to Zuma’s fall. Johnson, for

48 Ex parte Chairperson of the Constitutional Assembly: In Re Certification of the Constitution of the Republic of South Africa, 1996 [1996] ZACC 26, 1996 (4) SA 744 at para 111. The Court also said that ‘the separation of powers doctrine is … given expression in many different forms and made subject to checks and balances of different kinds.’ Ibid.
49 Ibid at 112.
53 I do not address questions concerning the role of the provinces under the Constitution. Cf J Wehner ‘Fiscal Federalism in South Africa’ (2000) 30 Publius 47.
54 Johnson (note 4 above) at 39.
55 Bracking (note 8 above) at 172.
example, explicitly links the Court’s ruling on the Public Protector – discussed in greater detail below— with the ANC’s subsequent defeats in local elections in Port Elizabeth, Johannesburg, and Pretoria.\footnote{Johnson (note 4 above) at 39.} Given that commentators have sometimes described the Court as otherwise aligned with many of the ANC’s policy preferences,\footnote{Roux (note 26 above) at 138.} its role here is all the more noteworthy.

A Democratic Alliance v Speaker of the National Assembly

The first case, Democratic Alliance v Speaker of the National Assembly, arose from the removal of members of the Economic Freedom Fighters party from the National Assembly’s chamber during President Zuma’s State of the Nation address on 12 February 2015.\footnote{‘South African Parliament Chaos as Malema MPs Heckle Zuma’ BBC News (12 February 2015).} The issue before the Court was the validity of Section 11 of the Powers, Privileges and Immunities of Parliament and Provincial Legislatures Act 4 of 2004. This section provided the statutory authority pursuant to which the Speaker of the National Assembly had ordered the parliamentarians’ removal. The petitioner in the case, a parliamentary party opposed to the ANC, argued that section 11 was a violation of the separation of powers.\footnote{Democratic Alliance v Speaker of the National Assembly & Others [2016] ZACC 8, 2016 2016 (3) SA 487 at para 7 (‘Democratic Alliance’).} Refusing to rely upon the Hyundai Court’s preference for reading legislation as in a manner consistent with constitutional dictates,\footnote{Investigating Directorate: Serious Economic Offences and Others v Hyundai Motor Distributors (Pty) Ltd & Others In re: Hyundai Motor Distributors (Pty) Ltd & Others v Smit NO & Others [2000] ZACC 12, 2000, 2001 (1) SA 545 (CC) at para 23.} Madlanga J held for a majority of the Court that Section 11 had been correctly read to allow the removal of members of the legislature. This broad reach, however, had a ‘chilling effect’ upon parliamentary free speech, and also ‘directly infringe[d] the immunities from criminal proceedings, arrest and imprisonment enjoyed by members’ pursuant to ss 58(1)(b) and 71(1)(b) of the Constitution.\footnote{Democratic Alliance (note 59 above) at paras 41 and 42.} Complementing this rights-based line of reasoning, the Court also offered a structural argument that took up the petitioner’s invitation to consider the separation of powers.\footnote{The Court, however, declined to consider or decide whether the challenged provision was ‘at variance with the separation of powers doctrine.’ Democratic Alliance (note 59 above) at para 53. Nevertheless, the structural implications of the decisions are sufficiently clear to be appropriately included here.} The aforementioned sections of the Constitution setting forth the privileges of parliamentarians, the Court explained, need not be read to allow for unlimited disruption by members intent on derailing legislative proceedings. But they did imply that the regulation of cameral speech had to be done by the relevant House itself, and not by an Act of Parliament.\footnote{Democratic Alliance (note 59 above) at para 47.} This reading, on my understanding, rested heavily on textual evidence rather than any functional analysis of allocating power at the cameral rather than the parliamentary level. On these grounds, the Court found that the February 2015 removal had been unlawful.

The Court could have taken a number of paths. Not only could it have reached a different result, but it could also have avoided the separation of powers issues raised and decided in Madlanga J’s majority opinion. For example, it might have accepted the Hyundai-based argument for construing s 11 narrowly. Jafta and Nkabinde JJ took this view in their
concurrency. This construction would have yielded a similar bottom line. It might even have left open the possibility of new legislation. (However, Jafta J’s opinion intimates that substitute legislation of greater clarity would be deemed ‘unconstitutional’.) Perhaps, as importantly, this narrow construction would also have vitiates the Court’s opportunity to make a broad statement on parliamentary free speech as well as the separation of powers. The opinion’s effects on elite and public opinion might, as a result, have been quite different.

On the merits, one can also easily imagine a different result. In respect to the separation of powers element of the case, indeed, it is noteworthy that American constitutional law has taken a different tack. The US Congress has enacted statutes that contain rules of cameral procedure since the 1790s, for instance to provide expedited consideration of trade treaties and otherwise to fashion frameworks for rapid legislative consideration of time-sensitive policy considerations. One legal scholar has developed a powerful argument for the position that Article I, Section 7 of the US Constitution, which endows each branch with power to set its own rules, renders such statute-based imposition of cameral rules unconstitutional, rather in the spirit of Democratic Alliance. Yet his position remains a minority one. The argument for permitting statutory rules of cameral procedure, moreover, is arguably stronger in South Africa than in the US. Under the American Constitution, both Houses and also the President normally play a role in the legislative process. Under Article 75 and related provisions of the South African Constitution, the president must generally assent, and has only a limited power to return laws. Hence, no argument can be proffered in terms of the South African Constitution (as is available under US Constitution) that the executive has shaped cameral rules. (That said, the effective fusion of the executive and the legislature in South Africa’s parliamentary system does allow the executive to influence the legislature’s agenda.)

Note also the distinctive way in which Democratic Alliance was a separation-of-powers decision. The decision allocates authority between different institutions created by the Constitution, but it does not move power between the branches. Rather, it shifts authority largely held by the parliament (with the exception of certain constraints in Article 79) to the level of the single chamber. The effects of this alternation of institutional relationships is ambiguous and complex. In the short term (i.e., the context of the particular decision), it benefited the plaintiffs – the Economic Freedom Fighters and the Democratic Alliance – who could not be sanctioned for past protests within the ANC-dominated legislature. In the longer term, however, the separation-of-powers element of the decision created no more than a temporary speed bump to the imposition of such punishments: the judgment did not make it more difficult to regulate parliamentary speech. To the contrary, all else being equal, it ought to be easier for a

---

64 Ibid at para 82.
65 I offer the US experience merely as a point of comparison; I do not mean to suggest that merely because the US courts have decided a question one way, that is the correct way to decide the issue.
68 Under Article 79, the President may refer a bill back to the National Assembly for reconsideration. If the reconsidered bill accommodates the President’s reservations, the President must assent to and sign the bill. If a reconsidered measure does not fully accommodate the President’s reservations, the President must either assent to and sign the Bill or refer it to the Constitutional Court for a decision on its constitutionality. If the Constitutional Court decides the bill is constitutional, the President must sign it.
69 The National Assembly’s power to promulgate internal rules, for examples, is described in Article 59(1)(b).
single chamber to enact (and also to repeal) rules limiting or punishing such speech than for the two chambers acting together to do so. In particular, where one party dominates a particular chamber, it seems likely that it will not be constrained in silencing the minority’s protests. The free speech component of Democratic Alliance, moreover, does not appear to impose much of a constraint on a cameral majority. The majority opinion here suggests that a chamber cannot ‘denude the privilege of its essential content’.

The precise effect of this ambiguous limiting language is hard to predict. Accordingly, it seems reasonable to infer that the short-term effect of Democratic Alliance – which was to affirm as lawful (and hence subtly to legitimate) a protest against President Zuma – may well be more important that its uncertain long-term implications.

### B Economic Freedom Fighters v Speaker of the National Assembly and Others

The second important decision of this period concerning corruption and state capture, Economic Freedom Fighter v Speaker of the National Assembly and Others, pertains to the Public Protector’s report on President Zuma’s spending on a private residence in Nkandla. In particular, it turns on the question of whether the Public Protector’s recommendations to the president and the National Assembly pursuant to s 182(1)(c) of the Constitution were binding or precatory. Stu Woolman has argued that Economic Freedom Fighters constitutes a critical ‘part of the Constitutional Court’s ongoing efforts to keep the train on the tracks – by ensuring that the state, specific state actors, and ‘well-connected’ private actors abide by the rule of law and are held accountable both to the Constitution and the people they serve.’ The question addressed here is precisely how the Court pursued these goals. After all, it is not entirely clear how ‘rule of law’ values inform a judicial decision as to which of a number of plausible readings of the law to adopt.

Nor is it self-evident how accountability should be defined. Mechanisms of accountability come in both legalistic and democratic forms. A surfeit of such mechanisms in a constitution can generate perverse and even undesirable outcomes. Both the vector and the appropriate quantum of accountability appropriate to preserve constitutional democracy must therefore be analysed with some precision.

The Nkandla residence of Jacob Zuma is located in KwaZulu Natal. It became an epicentre of public controversy as a result of a publicly-funded security upgrade ultimately costing some R246 million. The improvements to the property included a helipad, underground bunkers, facilities for security and their accommodation, a fire pool, a chicken-run, and fencing around

---

70 Democratic Alliance (note 59 above) at para 39. In the previous paragraph, the majority states that the privilege ‘can never go so far as to give members a licence so [as] to disrupt the proceedings of Parliament.’ Ibid at para 38. Hence, it appears that the Court envisages some constraint on parliamentary free speech as well as cameral power. Precisely how one would delineate the constraints in these two domains remains unclear.

71 [2016] ZACC 11, (3) SA 580 (CC).

72 S Woolman (note 32 above) at 174–175; S Woolman (note 32 above) at 159 (Woolman offers further reflection on that topic.)

73 The ‘rule of law’ can be defined in terms of constitutional law in a number of different ways; some are thin in the sense of providing little or no guidance as to the substantive content of the law. R Fallon Jr “The Rule of Law” as a Concept in Constitutional Discourse’ (1997) 97 Columbia Law Review 1.


the entire complex. In 2013, Public Protector Thuli Madonsela released a report on the Nkandla spending entitled ‘Secure in Comfort’ – the provisional, perhaps more telling, title had been ‘Opulence on a Grand Scale’. The report concluded that the improvements went well beyond the security-related alterations permitted by law; that Zuma had substantially benefited from the improvements; and that he was obligated to repay the state a reasonable amount of the resulting costs.\(^{76}\) Prior to the release of the Public Protector’s final report, the parliamentary ANC had also initiated its own investigation of the Nkandla spending. Rather unsurprisingly, this cleared Zuma of any impropriety.\(^{77}\) For a year after the Public Protector’s report had been submitted, a stalemate ensued. Neither the president nor Parliament took any steps to implement it. This stalling ultimately triggered applications by the Economic Freedom Fighters and Democratic Alliance parties seeking orders compelling those branches to take remedial actions, including determining the reasonable cost of the non-security measures and paying back to the state a reasonable percentage of that cost.\(^{78}\)

The Constitutional Court held first that it had jurisdiction over the complaints lodged by those two parties because they had alleged that both the President and the National Assembly had ‘breached their respective constitutional obligations’ in a matter of ‘high political importance.’\(^{79}\) The allegation that the President had personally failed to meet a constitutional obligation was pivotal to the Court’s analysis. In the course of this analysis, the Court again recognized that it was wading into ‘sensitive areas of separation of powers’.\(^{80}\) Indeed, the sensitivity of the underlying questions was in part a justification for its exercise of jurisdiction.

On the merits, the Court began by adumbrating the justifications for Chapter 9 institutions such as the Public Protector office. It described that office in particular as ‘invaluable’.\(^{81}\) To fulfil its role in combating corruption, graft, and other forms of impropriety, the Court explained, the Public Protector (like other Chapter 9 institutions) had constitutional guarantees of ‘independence, impartiality, dignity and effectiveness’.\(^{82}\) Mogoeng CJ then offered the simple yet devastating observation that unless the Public Protector possessed the power to

---

\(^{76}\) ‘Nkandla: It’s Still Opulence on A Grand Scale’ *Mail & Guardian* (20 March 2014).


\(^{78}\) A prior, but roughly contemporaneous, decision by the Supreme Court of Appeal held that the legislature, the executive and all organs of state were not merely under an obligation to respect the Public Protector’s rulings, but that they must view adverse findings, as well as recommendations and remedies, in reports by the Public Protector as having binding legal effect. *South African Broadcasting Corporation v Democratic Alliance* [2015] ZASCA 156, 2016 (2) SA 522 (SCA) at paras 47 and 54. The Supreme Court of Appeal’s conclusion flowed, in part, from its recognition that the Public Protector exists in order to ensure that the governors are held accountable to the same set of rules as the governed.

\(^{79}\) *Economic Freedom Fighters* (note 71 above) at para 24.

\(^{80}\) Ibid at para 19. The idea that the sensitivity of the issue and the personal involvement of the President created jurisdiction is almost the inverse of the standard view of jurisdictional constraints in American constitutional law. In US jurisprudence, the closer a question is to the apex of political power, and the more neuralgic it is, the less likely the US Supreme Court is to address the matter (or to evoke a form of deference that is functionally indistinguishable from the eschewal of jurisdiction). The US position is usually justified on grounds of expertise and democratic legitimacy. But neither reason is persuasive. There is no real difference between the questions of law and policy that arise when the Court declines jurisdiction and that arise in regulatory disputes which the Court routinely adjudicates. And it is implausible to think that democratic accountability will be effective when the executive controls the information available to the public about the success, and the costs, of a policy. A Huq ‘Structural Constitutionalism as Counter-Terrorism’ (2012) 100 *California Law Review* 887.

\(^{81}\) *Economic Freedom Fighters* (note 71 above) at para 52.

\(^{82}\) Ibid at para 50.
determine conclusively the appropriate remedies for official misconduct, ‘very few’ of the apex legislative and executive officials subject to her jurisdiction would ever comply with those recommendations. The mandatory force of her remedial conclusions, therefore, followed from the functional necessity of binding remedies in a context where compliance was typically incentive incompatible.

Turning to the \textit{legal} force of the Public Protector’s proposed remedies, the Court categorically rejected the notion that, in the absence of judicial review and constraint, they were advisory. A functional logic again played a key role here. As Mogoeng CJ explained: ‘One cannot talk about remedial action unless a remedy in the true sense is provided to address a complaint in a meaningful way.’ In rejecting the reasoning of an earlier High Court decision, Mogoeng CJ underscored the ‘irrationality’ of allowing alleged malefactors to decide on whether they should comply with the Public Protector’s recommendation. Such a posture, the Court explained, would be ‘at odds with the rule of law’. Its explanation for this last point, however, rested on the arguably circular assumption that the Public Protector’s recommendations constituted ‘law’ of the relevant kind.

The ensuing opinion did not endorse legislative (cameral) or executive authority. To the contrary, it affirmed the authority of a pivotal Chapter 9 Institution against the interests of both the legislative and the executive branch. This affirmation also came after the Public

\footnote{Ibid at paras 56 and 64. In subsequent sections of the opinion, Mogoeng CJ addressed and rejected the argument that ss 182(1) and 182(2) of the Constitution allowed the National Assembly to impose legislative restraints on the Public Protector’s actions. This is, however, in essence a response to a counter argument, rather than a positive justification for the Chapter 9 power at issue in the case.}

\footnote{A potential counterargument is that the Public Protector’s reports would be followed because of public pressure on the President. See E Posner & A Vermeule ‘The Credible Executive’ (2007) 74 University of Chicago Law Review 865. This is not, so to speak, a credible argument, even if it were not belied by the facts. Chief executives can often mitigate public pressure from accountability institutions by attacking those institutions as corrupt, or by ginning up political distractions.}

\footnote{\textit{Economic Freedom Fighters} (note 71 above) at paras 64, 65, and 68 (‘Remedial action must … be suitable and effective.’). Again, the disjunct between South African and American constitutional law is as striking as it is stark. In recent American constitutional jurisprudence, the supply of remedies for individual rights’ violations has gradually been constrained to the point where only the most gross and intentional transgressions of long-established law generate any kind of a judicial response. For an analysis of the relevant cases, and an exploration of the sociocultural context of these cases, see A Huq & G Lakier ‘Apparent Fault’ (2018) 131 Harvard Law Review 1525, and A Huq ‘Judicial Independence and the Rationing of Constitutional Remedies’ (2015) 65 Duke Law Journal 1. The US Supreme Court, however, has been far more willing to permit the litigants to invoke structural principles, even when doing so has no clear causal relationship to the values the Court purportedly wishes to vindicate. A Huq ‘Removal as a Political Question’ (2013) 65 Stanford Law Review 1. The net result of these postures is a highly regressive doctrinal structure.}

\footnote{\textit{Economic Freedom Fighters} (note 71 above) at para 72. Woolman notes that the Court was in fact arrogating to itself the power to determine what remedies pressed by Chapter 9 institutions are valid. Woolman (note 32 above) at 190. The decision hence might be read as an endorsement and enlargement of judicial power as opposed to a constraint on corruption.}

\footnote{The circularity arises because the Court assumes that the Public Protector’s recommended remedies are a ‘decision or step sanctioned as law’ that cannot be ‘ignored based purely on a contrary view’. \textit{Economic Freedom Fighters} (note 71 above) at para 75. This assertion assumes the legal force of those recommendations – which is precisely the legal question upon which the matter pivots. That said, the Court reasoning tracks the precedent and the structural analysis of the Constitution articulated by the Supreme Court of Appeal in \textit{South African Broadcasting Corporation v Democratic Alliance} [2015] ZASCA 156, 2016 (2) SA 522 (SCA): namely, that the Public Protector exists in order to ensure that the governors play by the same set of rules as the governed.}
Protector had published her bombshell ‘State of Capture’ report. It would be very surprising if the decision did not also influence both public and official perceptions of the latter document. Hence, like Democratic Alliance, the opinion in Economic Freedom Fighters likely had the immediate practical effect of advancing the political agenda of those who sought to constrain a hegemonic presidency. But unlike Democratic Alliance, the Economic Freedom Fighters decision rested more on a functional analysis of institutional powers than on a close reading of a specific constitutional text. (The contrast, to be sure, is one of degree and not absolute, since both opinions range over a variety of interpretive modes). The key term in Mogoeng CJ’s logic, as I read the case, concerned the (correct) recognition that officials fingered for graft by the public protector are likely to resist, or at least slow-walk in the face of those allegations. Nothing in the text of the Constitution conveys this assumption. Indeed, the textual recognition of broad legislative authority to regulate the Public Protector at least in some ways in sections 182(1) and 182(2) of the Constitution arguably cut in the other direction. Hence, Economic Freedom Fighters reflects a subtly different strategy of constitutional interpretation than Democratic Alliance. Had the latter’s textualism been applied exclusively, perhaps a different outcome would have ensued.

The opinions’ practical effect, moreover, diverge in an interesting way. In the Nkandla controversy, President Zuma had already agreed to the measures suggested by the Public Protector (to be sure under the shadow of constitutional litigation). Had the opinion come out the other way, it is at least possible that he would still have carried out these steps. Perhaps it would have been too embarrassing for him not to do so. Hence, it is at least possible to gloss the Court’s conclusion as merely confirming an existing political equilibrium. Even if one concludes that Zuma would have reneged on his earlier promise, though, the immediate effect of the opinion was relatively limited. The Nkandla graft, while not trivial, is hardly significant in comparison with the pattern of wholesale state capture associated with the Gupta family, a pattern that still imperils the nation’s fiscal standing. Rather, it seems more plausible to gloss its opinion in dynamic terms: elected actors, and in particular ANC officials, would know on a going-forward basis that they had a legal obligation of compliance with the Public Protector’s recommendations. That this ruling occurred in the context of the state capture controversy is a fact that should not be forgotten.

Consistent with this understanding, Ros Dixon and Theunis Roux argue that Economic Freedom Fighters embodied a precise, almost surgical, intervention that mitigated temporarily the systemic threat to democracy posed by state capture. They emphasize rightly that the Court alone would have insufficient authority to stem that threat completely. Rather, they describe the Court’s continued, if incremental, efficacy as a friction on democratic backsliding. Dixon and Roux underscore the long-term nature of the conflict between the ANC and the Court, praising the latter for its strategic foresight in placing out of bounds a strategy of noncompliance with Chapter 9 institutions. Dixon and Roux might have added that this insurance against lawlessness was not without its own flaws. If Public Protector reports cannot be ignored safely any more, the premium upon appointing captured loyalists to a leadership position in that

---

88 Economic Freedom Fighters (note 71 above) at para 100.

office rises. That is, even if the Court’s intervention limited state control, its effect probably had a shelf-life: its efficacy would evaporate when the hegemonic party captured the Public Protector itself.

C  United Democratic Movement v Speaker of the National Assembly

The third case of interest here arises out of parliamentary manoeuvring in response to President Zuma’s March 2017 decision to fire five cabinet members, including finance minister Pravin Gordhan and his deputy Mcebisi Jonas. The United Democratic Movement requested that the Speaker of the National Assembly permit a secret ballot on a motion of no confidence in Zuma. Its leader Bantu Holomisa argued that a secret ballot was appropriately used by the National Assembly to appoint the President. He also argued that ANC members would be subject to discipline and even expulsion if they did not support their President. In response, the Speaker concluded that she had no authority under the Constitution and the Rules of the National Assembly to permit a secret ballot. The United Democratic Movement, along with the Economic Freedom Fighters, the Inkatha Freedom Party and the Congress of the People, filed an application with the Constitutional Court seeking a declaration that the Speaker had erred, and that secret ballots were allowed under the Constitution in no-confidence motions.

The Court rejected the Speaker’s judgement that she lacked authority to conduct a secret ballot. It returned the matter to the Speaker to determine how to proceed in the light of that authority. Mogoeng CJ’s opinion begins with a strikingly long introductory essay on the separation of powers, and in particular on the role of Parliament as an institutional site for ‘accountability-ensuring mechanisms’. This introduction signalled the Court’s cognizance of the broader political context of its decision. The Court in particular honed in on the question of the efficacy of accountability mechanisms. Although it did not say that judgments of efficacy must account for contemporaneous political context, this assessment

---

90 Stu Woolman notes that the Court likely had in mind the demise of the Directorate of Special Operations, or Scorpions, which the government had eliminated in order to protect itself from an independent authority charged with investigating corruption both within the state and in society at large. S Woolman (note 32 above) at 175. See, especially, Glenister v President of the Republic of South Africa & Others [2001] ZACC 6, 2011 (3) SA 347 (CC)(Court holds that the government violated various provisions of the Constitution, read with international law, by eradicating an independent investigatory unit designed to combat fraud and bribery and replacing it with a body subject to direct executive oversight.)

91 N Onisha & S Chan ‘Firing of South Africa’s Finance Minister Widens a Political Rift’ The New York Times (31 March 2017). This decision – following prior attempts to manipulate the Finance Ministry by the firing of Nhlanhla Nene in December 2015 – played some role in the ultimate downgrading of South African sovereign debt by two rating agencies.

92 M Gallens ‘Secret Ballot in Vote of No Confidence Will Protect ANC MPs – Says UDM’ News24 (April 10, 2017). Many ANC senior figures, including Ramaphosa, had criticized the firings. Onisha & Chan (note 91 above). It is difficult to know, however, whether rank-and-file members do vote freely in a more consequential no-confidence motion.


94 Ibid at paras 1 to 10. In para 33, the Court reiterated the theme of accountability and the ‘reality that constitutional office-bearers occupy their positions of authority on behalf of and for the common good of people.’ The very repetition of this (by then plainly) counterfactual ‘reality’ works to underscore the Court’s cognizance of the contemporary political conjuncture.

95 Ibid at para 12. The theme is repeated later in the judgment: it suggests a constitutional obligation to have ‘effective mechanisms in place’ to hold ‘Members of the Executive accountable’. Ibid at para 40.
seems implicit in its discussion. A contextual approach to efficacy was also implicit in its extended discussion of no-confidence motions under s 102 of the Constitution. It characterized these motions as a ‘crucial consequence-management or good-governance’ issue necessary to ensure the effectiveness of existing mechanisms, in-between the general elections to articulate ‘dissatisfaction’ with executive conduct.96 The Court, in short, organized its analysis around the pragmatic question of efficacy in a way that invites, and perhaps even compels, a contextual and situated judgment of contemporaneous political circumstances.

All this threshold analysis, however, had strikingly little to do with the actual holding. The decision turned, as in Democratic Alliance, on a relatively simple point of textual analysis. Section 57(1) gives the National Assembly power to ‘make rules and orders concerning its business’ and does not mention either secret or open balloting, the Court observed. Therefore, the Constitution permitted both secret and open balloting, hence the Speaker’s assumption of no authority.97 Although the Court offered some comments on the benefits of a secret ballot in the context of a no-confidence motion, it is hard to read these stray thoughts as anything more than supplementing its central, textual ground of decision.98 This textual reading, it should be underscored, could have been reached without any of the anticipatory paraphernalia of the decision. The result is an opinion with a slightly bathetic note.

The principal effect of United Democratic Movement, like that of Democratic Alliance, quickly materialized. A vote of no-confidence using a secret ballot, which Zuma survived, was ultimately held in August 2017.99 Even if that vote was not decisive, Zuma and his allies had struggled mightily to avoid it. The narrowness of the vote margin, moreover, demonstrated the extent to which Zuma and his allies had lost support within and control over the ANC. That information, aired publicly by the vote, likely weakened Zuma further because it increased the likelihood that wavering ANC members might later ‘defect’ from his caucus, and openly resist his presidency. Paradoxically, therefore, the secrecy of the no-confidence vote allowed legislators to generate a credible public signal of the extent of dissatisfaction with the Zuma presidency in a way that lowered the anticipated costs of defecting from that regime subsequently. It thus anticipated, and rendered more likely, Zuma’s ultimate February 2018 ouster.

The decision will also have some enduring consequences. In the longer term, the decision means that a dominant majority party such as the ANC cannot use the veil of parliamentary secrecy to shield, and so ignore, fissures within its own coalition as a way to prevent the ouster of incumbent ministers or presidents. The effect of the decision, therefore, is not so much to shift power from one branch to another. It is instead to reallocate authority from the party in government to the more diffuse parliamentary party. In the context of a ‘dominant party’ system of a kind that has become prevalent across sub-Saharan Africa,100 this shuffling of intraparty

96 Ibid at para 47.
97 Ibid at paras 58 to 61. The Court further held in para 67 that the Speaker had erred in reading the relevant National Assembly rules not to allow for secret balloting.
98 Ibid at paras 73 to 76.
99 S Allison ‘Jacob Zuma Narrowly Survives No-Confidence Motion in South African Parliament’ The Guardian (9 August 2017). However, had Zuma been truly concerned about his immediate survival, he would have likely applied some pressure on the Speaker.
100 There are a number of different definitions of ‘dominant party’ system that have been applied in the Sub-Saharan African context, although the ANC fits all of them. For a survey of the relevant work, see M Bogaards ‘Counting Parties and Identifying Dominant Party Systems in Africa’ (2004) 43 European Journal of Political Research 173.
power – with its concomitant implication of a less concentrated form of parastatal political power – is perhaps a more deeply significant institutional change than is readily apparent.\(^{101}\)

One way to theorize *United Democratic Movement* is in light of the ‘focal point’ account of constitutional enforcement developed by the political scientist Barry Weingast and others.\(^{102}\) On Weingast’s account, constitutional rules provide crisp and readily available focal points that act as a signal to diffuse, uncoordinated social and political actors that democratic norms are imperiled. The absence of a focal point will render popular and oppositional resistance to the antidemocratic consolidation of political power more costly to organize and hence less effective. Weingast’s theory concerns the content of constitutional rules themselves, rather than their implications for parliamentary voting rules. But it can be extended to cover the logic of *United Democratic Movement*: By facilitating a motion of no-confidence, that is, the Court enabled opponents of an entrenched presidential clique to discover their own numerosity. By facilitating the disclosure of this information, the decision lowered epistemic barriers to increased coordination among political opponents of an incumbent president to check his or her power.

**D Corruption Watch NPC v President of the Republic of South Africa**

One final decision is worth mentioning even though it was heard some two weeks after Zuma had resigned and was not handed down for another six months. *Corruption Watch* concerned Zuma’s efforts while in office to remove former National Director of Public Prosecutions Mxolisi Nxasana by offering him a large financial settlement conditional on his departure from the position, and to install in his place Shaun Abrahams.\(^{103}\) Because the case arose out of Zuma’s effort to obtain greater control over accountability mechanisms, it is related to the ongoing process of state capture. The decision also raises the same question of how to craft institutional rules to respond to the prospect of state capture. Hence, it makes sense to treat the case as part of the same jurisprudence.

The language and result of the case also betray a concern with state capture. The Court held that s 179(4) of the Constitution guarantees the independence of the prosecuting authority as a safeguard against the risk of political capture by ‘criminals … holding positions of influence’.\(^{104}\) Concluding that Zuma had sought to ‘get rid of Mr. Nxasana at all costs’, Madlanga J explained for the Court that ‘effectively buying out’ the Director of Public Prosecutions constituted an improper infringement on that office’s independence.\(^{105}\) He further held that Abrahams’s appointment was as a result ‘constitutionally invalid’.\(^{106}\)

*Corruption Watch* is the only one of these four decisions that deals directly with the authority of the President under the Constitution. It sharply narrows that authority in the cause of democracy. Its reasoning sounds in the same functionalist and consequentialist register as *United Democratic Movement*. Indeed, it is not impossible to imagine judges reasoning that Nxasana’s


\(^{103}\) *Corruption Watch NPC and Others v President of the Republic of South Africa and Others; Nxasana v Corruption Watch NPC and Others* [2018] ZACC 23, 2018 (10) BCLR 1179 (CC), 2018 (2) SACR 442 (CC).

\(^{104}\) *Corruption Watch NPC* (note 103 above) at para 19.

\(^{105}\) *Corruption Watch NPC* (note 103 above) at paras 26 to 29.

\(^{106}\) *Corruption Watch NPC* (note 103 above) at para 35.
decision to accept the financial offer from Zuma meant that his departure could not be construed as a termination, but rather as a voluntary action. The Court (correctly, in my view) recognized that the course of dealing between Zuma and Nxasana was salient to an understanding of the latter’s final decision, and declined to analyse their deal formalistically. But this tactical and fruitful form of analysis cannot be taken for granted. Again, the US courts’ tendency to view the question of removal authority in formalist terms stands in telling contrast.  

Given the timing of the hearing and ultimate resolution of the case, the decision could have had no direct impact on Zuma’s continued ability to hold and to wield power. Indeed, it is reasonable to speculate that Zuma’s fall provided a formative backdrop to the Court’s deliberations. Rather, the effect of Corruption Watch will be felt during later presidencies. Subsequent holders of the office will lack an important tool for controlling the progress of criminal investigations of their own conduct, or that of close allies.

E The tactical challenge of preserving constitutional democracy

The four decisions discussed in this part all concerned the Constitutional Court’s response to the problem of state capture. The cases respond in part to the pressure imposed on the value of ‘legality’; the latter can be understood as simple compliance with constraints on the behaviour of senior executive branch officials. In addition, they can be glossed in terms of a background worry about the sound operation of ‘democracy’; the latter can be understood as the ability of various democratically chosen actors, including members of the legislative minority, to air complaints about government malfeasance and wrong-doing as a necessary epistemic backdrop to renewed democratic contestation. In their different ways, therefore, all four cases concern the problem of how a court can protect constitutional democracy, or what Ginsburg and I have called liberal constitutional democracy, from the potentially disabling shadow of state capture.

If the opinions are united by their cognizance of that shadow, and their awareness of the political realities in which the Court operated, they are also marked by a diversity of analytic methods and theoretical commitments. One way of understanding this diversity – and of linking it to the underlying theme of judicial responsibility for constitutional democracy in the teeth of state capture – is to see that judicial task in the face of democratic backsliding is not a matter of defending abstract principles. It is as a matter of tactics. The Court in these four cases was not following the unfolding logic of a single structural principle or textual command. It was, rather, engaged in a highly contextual, situationally dependent exercise by responding to the many and various ways in which an antidemocratic movement or faction was seeking to stymie institutional levers of legality or democracy. The ways in which democracy can be undermined are plural and unpredictable in their forms and sequencing. As a result, a court trying to maintain constitutional democracy cannot be theoretically rigid, or attend only to one element of institutional design. It must rather be tactical, responding in a pragmatic fashion to the particulars of the situation, and not bound to an a priori constitutional theory. Of course, the tolerance for such tactical thinking may vary between different national and socioeconomic contexts. As a pragmatic matter, courts inclined to perform a systemically supportive role may be well advised to cultivate a tolerance for such action.

---

107 Huq ‘Removal as a Political Question’ (note 85 above).
108 I explore this theme at greater length in Ginsburg & Huq How to Save a Constitutional Democracy (note 27 above).
To appreciate the extent to which these decisions are internally heterogeneous, consider three different margins of comparison: the Court’s choice of interpretive methodologies; the decisions’ effect on the distribution of governmental power across institutions; and the time-frame (long-term versus short-term) of their main effect. First, a range of different methodological approaches are visible across the four decisions discussed in this part. For example, Democratic Alliance and United Democratic Movement are, at their core, relatively simple glosses upon the constitutional text. Consequentialist analysis is largely absent in Democratic Alliance, while it is present but apparently superfluous in United Democratic Movement. In contrast, Corruption Watch starts from a highly purposive understanding of the Constitution’s protection of prosecutorial independence. It generated from that more holistic understanding an inference about the scope of presidential authority to reach side-deals with senior officials. That inference, however, is hard to extract directly from the relevant constitutional text without a dose of contextualizing, pragmatic reasoning. Similarly, the result in Economic Freedom Fighters flowed from an understanding of the likely incentives of institutional actors faced with corruption and graft investigations, rather than following narrowly from the constitutional text. Accordingly, the decisions evince a mixture of methodological commitments. Indeed, they sometimes tacked between those commitments in the context of a single decision. It is thus hard to see any singular ‘theory’ of the separation of powers at work across these cases. The Constitutional Court was not, in other words, reasoning from a priori, abstract grounds of constitutional theory to specific outcome in these cases.

Second, the decisions did not have parallel effects on institutional power in the sense of uniformly decreasing executive power or increasing legislative authority. Rather, they had dispersed and distinct institutional logics. At the same time, it is likely that their logics all proved conducive in either the immediate or the long term to limitations on state capture. Economic Freedom Fighters and Corruption Watch hence strengthened the hand of accountability institutions. The former concerned a Chapter 9 body, while the latter concerned an entity within the executive. In contrast, Democratic Alliance and United Democratic Movement did not concern directly the distribution of authority between branches. Rather, they both altered the distribution of effectual authority within the National Assembly. Moreover, in my view, United Democratic Movement is best understood as a case concerning the distribution of power within a dominant party (which often operated as an adjunct to the constitutional state) rather than a decision about legislative authority per se. It thus demonstrates the Court’s awareness of the interaction between constitutional structures and political-party processes.109 It also evinces a measure of sophistication in understanding how best to shape the latter. It is striking to note that only one decision (Corruption Watch) directly concerned the scope of presidential power even though all these decisions are plainly preoccupied with the problem of restraining a seemingly lawless presidency.

These decisions, finally, operated in different timeframes. The immediate effect of Economic Freedom Fighters, for example, may have been dampened by the fact that President Zuma had ostensibly agreed to follow the remedies identified by the Public Protector – though the decision clearly reverberated through the general public and ANC rank and file members. Its effect was also long-term, insofar as it insulated and strengthened the hand of Chapter 9

109 The insight is all the more striking insofar as most post-1945 constitutions contain only ‘a brief and general article (or articles) acknowledging the role of political parties in contributing to the formation of the public will in a democracy.’ Gardbaum (note 101 above) at 241.
institutions in future fights with the president. In contrast, the effect of Democratic Alliance was likely limited to the short-term because it made it easier (rather than more difficult) for a dominant party to punish or to exclude protesting minority or opposition party members from the National Assembly. In the long term, this effect would fade or even reverse. Similarly, in United Democratic Movement, there was a quite clear immediate effect in the form of the August 2017 secret ballot on a no-confidence motion lodged against President Zuma. But that decision will likely also continue to shape the ability of a dominant party to exercise control over the legislative process so as to protect an incumbent president from various forms of accountability. Like Economic Freedom Fighters, therefore, its effects are temporally plural and complex.

In my view, the best way to read these opinions is in terms of the Constitutional Court’s conscious assumption of a role defending what Mogoeng CJ and Madlanga J called ‘constitutional democracy.’ On the ground, this defence turns out to be not a matter of applying some a priori theory of institutional relations under a constitution. Nor is it a matter of defending one particular institution’s prerogatives from the threat of another. Rather, it is a pragmatic, context-sensitive exercise in responding to the particular threats presented to legality and democracy as systemic values. Fifteen years ago, Theunis Roux demonstrated that the Court was ‘adept’ at a ‘kind of strategic behaviour.’ Albeit in relation to quite distinct issues, this form of pragmatism still seems to be alive today. While it would be premature to suggest that the Court has ‘succeeded’ tout court, the evidence certainly suggests that this tactical approach to the separation of powers can play an important, if partial, role in preventing a gradual process of erosion in the quality of constitutional democracy in the context of pervasive, and pernicious, state capture.

IV CONCLUSION: IMPLICATIONS FOR SEPARATION OF POWERS THEORY

The Constitution, as glossed by the Constitutional Court, reflects a commitment to the separation of powers that works cunningly to protect the Founding Provision’s guarantee of a government designed ‘to ensure accountability, responsiveness and openness’. These goals of constitutional design map closely onto the purposes of the separation of powers in other constitutional systems. It is therefore reasonable to consider the South African case law as a lens to consider the way in which different articulations (or conceptions) of the separation of powers can be implemented. The latter question has special resonance for American constitutionalism, where profound disagreements on fundamentals still characterize scholarly and jurisprudential conversations on the separation of powers. In concluding, therefore, I draw upon the South African jurisprudence mapped in Part III to critique American theories of the separation of powers.

In the American context, the dominant approach to the separation of powers (particularly among the Justices of the US Supreme Court) is a formalist one. This approach understands each branch as a distinctive and stand-alone entity with a lexically defined set of powers, usually

---

110 Roux (note 25 above) at 93.
113 The debates are canvassed and dissected in A Huq ‘Separation of Powers Metatheory’ (2018) 118 Columbia Law Review 1517, from which the following descriptions are drawn.
derived from textual exegesis or inference of the original meaning of the first three Articles of the Constitution. The result is a judicial effort to discern answers to unexpected institutional design questions in the inchoate and murky waters of late eighteenth-century constitutional debates. Perhaps the leading (or at least the most sophisticated) scholarly criticism points out that it is often not possible to identify ex ante a specific government action as legislative, executive, or judicial because there is commonly an observational equivalence between those forms of state action. Scholars working in this vein are sceptical of any approach that treats the ‘relevant constitutional language … as a set of descriptive labels, a set of terms like “executive”, “state”, or “judicial” (terms that seem ripe for definition or drawing boundaries), [such that] texts are then matched against the challenged practice under review’. They are sceptical that the ‘branch’ is the truly relevant unit of analysis if one wishes to understand and predict the actions of official actors.

The South African experience suggests that both formalism as a separation-of-powers strategy, and the highly sceptical response to formalism, are both flawed. The Constitutional Court has worked with reasonable success to preserve constitutional democracy as a going concern by tacking between different methodological registers rather than allowing the text or some sort of ‘original understanding’ to dominate its reasoning. The success of this judicial enterprise has two implications for the American debate. First, it suggests that formalism’s rigid focus on the three constitutional branches defined a priori in relation to original understanding may well have perversely destabilizing consequences. It is hard to see how formalist reasoning could conduce to the kind of thoughtfully tactical, and hence effective, interventions as those mapped in Part III. In addition, the Constitutional Court’s willingness to account for the rule of intra-branch and external actors, such as political parties, starkly contrasts with the myopic assumption of American formalism that analysis must proceed in a rigidly atomistic manner, in which branches provide the basic and indissoluble blocks of constitutional theory. Separation-of-powers formalism, in short, would fail the task taken on by the Constitutional Court.

At the same time, the Constitutional Court’s response to the Zuma presidency suggests that it would be premature to lapse into unmitigated scepticism about the separation of powers as a framework for analysing the production of systemic public goods such as the rule of law and democracy. Such scepticism is belied by the ability of the Constitutional Court to identify legal interventions that serve a useful purpose, and then to craft decisions that select for immediate and long-term effects in ways that are plausibly responsive to the project of protecting constitutional democracy.

The cases canvassed above underscore a different, more useful way of thinking about the purposes of governmental design decisions. They suggest that to the extent that the separation of powers is looked to as safeguard of constitutional democracy, it might be best comprehended as a pool of institutional resources to be leveraged in the face of mutative challenges to that order. That pool is institutionally diverse. It encompasses not just the three branches but also Chapter 9 institutions, the Director of Public Prosecutions and organizations that cut across all branches and organs of state, such as political parties. Institutional pluralism of this sort gives the Court a

114 The standard citation for an academic defence of this theory is S Calabresi & S Prakash ‘The President’s Power to Execute the Laws’ (1994) 104 Yale Law Journal 541.
range of options for responding to diverse forms of democratic backsliding: different redoubts of resistance to democratic backsliding can be protected. This matters because modern challenges to constitutional democracy tend to be opportunistic and unpredictable in their approaches. Attacks on democracy, that is, require no robust theory to be effective. Instead, these challenges take advantage of whatever weaknesses exist in a given constitutional structure.

Accordingly, the judicial enforcement of the separation of powers to promote constitutional democracy must also be agile, responsive, and provisional, rather than a matter of static relations calibrated by a priori theory, therefore, the separation of powers doctrine should be comprehended as a fluid (even open) set of concepts and potential judicial moves. Contra the sceptics, this project can bear fruit by enlarging the space for democracy’s allies (often in the parliamentary opposition, and frequently mobilized by a fear of being excluded permanently from power). Even if courts cannot save democracy on their own, that is, they can create political space in which others can do so. The greater the ensuing flexibility, the greater the chance that a court sympathetic to democratic norms might find ways to delay and defer state capture long enough for popular mobilization to neutralize more directly that threat. If there is no guarantee of success, recent South African experience nonetheless suggests the project is hardly one that is doomed to failure.

To be clear, I do not think that a High Court will necessarily be committed to democracy, or will be willing to take risks to defend democracy. A Huq ‘Democratic Erosion and the Courts: Comparative Perspectives’ (2018) 93 New York University Law Review Online 21. In the United States, for example, the Supreme Court as currently constituted is probably more hostile than friendly to democratic norms, specifically through its willingness to enable and even accelerate concentrations of executive and corporate power; its unwillingness to tolerate campaign finance reform; its openness to restrictions on the franchise and self-dealing by political elites in the form of partisan gerrymandering; its willingness to allow aligned political elites to transform policy disputes into novel constitutional problems (eg, the challenges to the Affordable Care Act); and its reluctance to rein in police and security services who overstep their public safety mission, and instead curb free speech or assembly rights. One of the open questions in American constitutionalism is the extent to which democratic majorities – if allowed to gain power – will tolerate such an antidemocratic and even unjust institutions.

---

\(^{117}\) To be clear, I do not think that a High Court will necessarily be committed to democracy, or will be willing to take risks to defend democracy. A Huq ‘Democratic Erosion and the Courts: Comparative Perspectives’ (2018) 93 New York University Law Review Online 21. In the United States, for example, the Supreme Court as currently constituted is probably more hostile than friendly to democratic norms, specifically through its willingness to enable and even accelerate concentrations of executive and corporate power; its unwillingness to tolerate campaign finance reform; its openness to restrictions on the franchise and self-dealing by political elites in the form of partisan gerrymandering; its willingness to allow aligned political elites to transform policy disputes into novel constitutional problems (eg, the challenges to the Affordable Care Act); and its reluctance to rein in police and security services who overstep their public safety mission, and instead curb free speech or assembly rights. One of the open questions in American constitutionalism is the extent to which democratic majorities – if allowed to gain power – will tolerate such an antidemocratic and even unjust institutions.
Precautionary Constitutionalism, Representative Democracy and Political Corruption

FIROZ CACHALIA

ABSTRACT: The Constitutional Court has over the past several years intervened in a significant number of cases that set concrete rules for the manner in which the internal procedures of our representative branches should operate. Once one overcomes anodyne expressions of the separation of powers doctrine, the departure from ‘well established’ understandings of the proper relationship between the judiciary and other branches in constitutional democracies begins to make sense in South Africa. That said, our fundamental commitment to representative democracy and a ‘separation of powers’ should urge some caution when it comes to judicial intervention in the ‘internal functioning’ of the representative branches. This article has put forward a ‘democratic process’ theory, which includes a specific case for ‘representation-reinforcing review’ aimed at promoting ‘principal/agent’ accountability. Given the deep commitment of the basic law to self-government in multiple forms, this theory holds that where self-dealing representatives: (a) fail to fulfil their custodial responsibilities for public resources on behalf of ‘the people’; (b) engage in self-dealing; (c) are ‘captured’ by private interests; or (d) create corrupt patronage systems that threaten the entire democratic project, then the judiciary is well within its rights to impose binding legal constraints on the political process in order to secure accountability. Of course, where accountability ends, and meddling starts, can sometimes be difficult to ascertain. On several occasions the Court has failed to offer a persuasive argument for intervention in the internal processes of the representative branches of government. But while the outcomes might appear jarring on a first read, this article shows that the Court’s implicit recognition in recent cases of constitutional standards governing the relationship between representatives and voters (a ‘principal/agent’ relationship) in a representative democracy and the express principle of accountability in a constitutional state provides justification for the more troubling decisions reached and bright lines for future cases. The case for democracy-reinforcing review propounded here enables us to ascertain, as best as we can, when the rough and tumble of democratic politics – as a deep constitutional commitment – must be respected and when the democratic process has become so utterly dysfunctional and our representatives so remiss in the observance of their constitutional obligation of accountability to ‘the people’, that it renders our representative democracy ‘representative’ and ‘democratic’ in name only.

KEYWORDS: representative democracy, accountability, democracy-reinforcing process theory, corruption

AUTHOR: Professor of Law, University of the Witwatersrand.
Email: Firoz.Cachalia@wits.ac.za
ACKNOWLEDGEMENTS: This article owes a significant debt to my collaboration with Stu Woolman over the past year and to conversations that we have had over a longer stretch of time. He has read multiple iterations of this work and then edited this version from first word to last, so that, the theses put forward are tighter, stronger, better grounded in the law and secondary literature, and the whole is now leaner and, I think, more coherent. I owe him a huge debt. I would also like to thank Justice Albie Sachs for consistently finding the time to read my work, as well as my referees, Professors Klaaren and Pieterse for reading this article at various stages and enabling me to rethink and sharpen the arguments.
The day might come when the leader says that to have a gold bed or diamond studded toilet paper holder is to support local industry, but I rather doubt it. Our corruption, should it come, would be of the more sophisticated kind, as befits a developed country. … Constitutions both express and tame power. They are built not on trust but on mistrust, and not just of the other side, but of ourselves.1 — Albie Sachs

A dependence upon the people, is no doubt, the primary control on the government; but experience has taught mankind the necessity of auxiliary precautions. — James Madison

I INTRODUCTION

A Precautionary Constitutionalism2

South Africa had its founding ‘constitutional moment’ just more than two decades ago, putting an end to the political and legal institutions of the apartheid system and inaugurating a new political and legal order based on democracy, human rights and the rule of law. This specific transition appeared to be part of a global and irreversible trend towards democratic systems and constitutionalism.3 Political transition to a non-racial constitutional order through deliberative acts of bargaining and political choice, was a significant achievement in a society with a history of colonial conquest and despoliation. South Africa’s judiciary supported and legitimated this transition4 and over the next two decades developed an impressive body of ‘aspirational jurisprudence’ concerned both with the enforcement of individual rights and the rule of law. Many judges and constitutional scholars saw the Constitution of the Republic of South

2 M Dorf ‘The Aspirational Constitution’ (2009) 77 George Washington Law Review 1631. Constitutions are aspirational in the sense that they identify and symbolise the shared ideals of a political community and the goals to be achieved over time. The idea, for instance, that the Constitution has transformative goals is widely accepted in the legal community and by judges. But constitutions are also precautionary devices that aim at managing the risks of ‘backsliding’ and even failure associated with democratic politics and the abusive or corrupt exercise of power. The South African Constitution is no different. See also J Tullis & S Macebo (eds) The Limits of Constitutional Democracy (2010); T Ginsburg & AZ Huq How To Save Constitutional Democracy (2019). Both books offer analysis of the kinds of factors that impair the functioning of constitutional democracies.
Africa, 1996 (Constitution) as a blueprint for ‘transformation’ and social change. All that such a transformation required was a change in legal culture from formalism to a substantive commitment to social democracy, and a good-faith implementation of the Constitution’s immanent normative postulates by the political branches.

Absent in this political-constitutional moment was a realistic conception of the risks inherent in democratic politics and of forms of political dysfunction that can threaten the practice of democracy and the prospects of ‘transformation’. Sam Issacharoff, in these pages, pointed to the need for the Court to ‘articulate a theory of proper democratic politics in order to discharge a principled role in engaging political process failures’. Such risks, leading potentially to forms of ‘dysfunction’ that need a judicial remedy, can take various forms; abuses of power, the undue influence of monied elites, majoritarian or minoritarian oppression, self-dealing by incumbents, authoritarianism, a


6  Precautionary constitutionalism is concerned with ‘risk mitigation’ both ex ante and ex post. But in this article, my focus is on the latter, ie on forms of ‘political dysfunction’ that have become manifest in the functioning of democratic institutions in ‘post-apartheid’ South Africa.

7  S Issacharoff ‘The Democratic Risk to Democratic Transitions’ (2016) 8 Constitutional Court Review 97. My case for the expansion of judicial review to consider questions concerning the ‘internal functioning’ of representative bodies generally considered to be non-reviewable on a traditional separation of powers grounds, is premised on a recognition that the scope of review in any particular jurisdiction is in part a contextual question, which depends on whether certain background conditions exist, in particular whether a country has good, working democratic institutions, and whether the implicit norms that support democratic practices are generally observed, especially by those who hold elective office.

lack of accountability,9 executive dominance of legislatures,10 party dominance,11 various forms of oligarchic12 and plutocratic capture,13 unjust socio-economic inequalities,14 and political corruption
Constitutional Court Review 2019

– the particular focus of this article. Constitutions should therefore be understood, not only aspirationally as texts which articulate a conception of justice, but as precautionary devices for managing risks associated with the exercise of power in a representative democracy. At the ‘higher law’ moment of its formulation and founding, the ‘risk mitigation’ question arises as a question primarily of institutional design. Once vices inherent in ‘ordinary politics’ and the ‘rule of men’ arise, the question that follows is how the extant constitution ought to be interpreted in the light of political ‘dysfunction’ that has become manifest.

Which political risks and forms of ‘dysfunction’ require a constitutional remedy is a question that depends on alternative theoretical accounts of the constitutional project and democratic ideals. But democratic constitutions always have a dual purpose: facilitating the exercise of ‘good power’ and constraining the exercise of ‘bad power’. Protecting the integrity of the

15 ‘Political corruption’ as I use the concept, is not limited to ‘quid pro quo’ corruption involving an improper exchange of inducements and benefits between a private party and a public official exercising public functions. This kind of corruption is a criminal offence, and is dealt with through the criminal justice system. See the Prevention of Corrupt Activities Act 12 of 2004. But political corruption is a broader concept. It includes corrupt, ‘self-dealing’ by elected public representatives that threatens the functioning of the democratic institutions of ‘republican self-government’. Citizens United v FEC 558 US 310 (2010). Justice Steven’s wrote: ‘When they bought our constitution into being, the Framers had their minds trained on a threat to republican self-government’, in the form of political corruption. Ibid at 452. It is for this reason that constitutions provide for exceptional remedies through the political process, like impeachment, which are not dependent on proof of criminal conduct. Nor is political corruption limited to discreet acts. It includes institutional and systemic corruption. See Lawrence Lessig ‘On the Legitimate Aim of Congressional of Regulation of Political Speech: An Originalist View’ in LC Bollinger & GR Stone The Free Speech Century (2019) at p.99. That is how ‘state capture’ in South Africa should be understood. This conceptualisation explains why political corruption properly raises constitutionally salient questions and informs my analysis of recent cases of ‘unorthodox’ intervention by the judiciary in the political processes of the National Assembly. Theunis Roux has suggested that extant corruption is an ‘epiphenomenon’ with structural roots – while Stu Woolman has noted that it can also be traced to a host of ‘wicked problems’ beyond South African borders and local sovereign control. If so, then judicial review and constitutions can provide only limited remedies. They may be correct. However, the control of abuses of political power, including corrupt exercises of such power has always provided one of the more important justifications for constitutional constraints in a democracy.

16 A Vermeule ‘Precautionary Principles in Constitutional Law’ Harvard Public Law Working Paper 11–20 (2012), available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1930427. The precautionary principle in constitutional law covers prophylactic measures taken in ‘designing constitutions’ [the drafting stage] and once they are operational, [the implementation stage] measures, including ones enforced by the judiciary, to manage actual risks as well as speculative ones, which if actualised, have devastating consequences. This perspective informs my suggestion that the Constitution should be read as instantiating an ‘anti-corruption’ principle, and can plausibly be so read.

17 J Elster Ulysses and the Sirens: Studies in Rationality and Irrationality (1979); J Elster Securities Against Misrule: Juries, Assemblies, Elections (2013). Elster argues that the primary task of constitutional design is a negative one of weeding out self-interest, passion, prejudice and bias, rather than a positive one of producing good outcomes. He makes the argument that written constitutions are needed to constrain legislatures ‘in order to prevent those in power from using their power to keep their power’. Ibid at 191–192. We have learnt from experience that ‘self-dealing’ elected representatives can misuse otherwise legitimate power, to evade accountability and entrenched themselves in power. In such circumstances, judicially enforced constitutional constraints can prove to be important in securing at least some of the conditions for the preservation of a project of democratic constitutionalism.

18 B Ackerman We the People, Volume 1: Foundations (1991)(Ackerman draws a distinction between ‘constitutional politics’ and ‘ordinary politics’)

institutions of self-government and the democratic process from the negative consequences of improper influence and the corrupt exercise of power provides one of the basic premises of precautionary constitutionalism.

B State capture and the Constitution

‘State capture’, a form of political corruption, entered our public lexicon after the publication of two reports by the Public Protector. The first report found that public funds had been spent on the President’s private homestead for non-security purposes.\(^20\) The second, that he had effectively handed over his constitutional authority to appoint Cabinet Ministers to private individuals who are in business with his son.\(^21\) A subsequent report by a group of academics came to the conclusion that under the Zuma administration, public power was effectively being exercised by a ‘shadow state’ – a well-organised clientelistic and patronage network.\(^22\) Despite the overwhelming evidence of systemic corruption, the National Assembly – the body ‘elected to represent the people and to ensure government by the people under the constitution’\(^23\) – failed to discharge its constitutional obligation to hold the President to account.\(^24\)

‘State capture’ raises fundamental questions both of constitutional integrity and democratic legitimacy since the normative democratic theory underpinning the Constitution envisages a political process that is sufficiently under the control of both the electorate and their representatives.\(^25\) When a ‘faction’ of the state, acting in concert with a network of unauthorised private actors, usurps the rights and powers of the popular sovereign, the representative mechanism is no longer functioning as envisaged by the Constitution.

How should constitutional theory and law respond to problems of ‘political dysfunction’ in the form of political corruption and state capture? What is the role of judicial review in

---


\(^{21}\) Office of the Public Protector State of Capture: Report on an Investigation into Alleged Improper and Unethical Conduct by the President & Other State Functionaries Relating to the Alleged Improper Relationships and Involvement of the Gupta Family in the Removal and Appointment of Ministers and Directors of State Owned Enterprises Resulting in Improper and Possibly Corrupt Award of State Contracts and Benefits to the Gupta Family Business (Report 6 of 2016/2017) (‘State of Capture’).


\(^{23}\) Constitution s 42(3).

\(^{24}\) Ibid s 55(2). Accountability is also one of our founding constitutional values. See Constitution s 1.

\(^{25}\) Vermeule (note 16 above). Adrian Vermeule characterises this kind of ‘dysfunction’, where representatives are insufficiently constrained by voters, as a kind of ‘agency slack’, ie, the defect is in the presumed relationship between voters and their representatives. The identification of a principal/agent problem in a representative government can also be found in John Stuart Mill’s essay ‘On Liberty’: ‘In time … elective government … became subject to the observations and criticisms which wait upon a great existing fact. It is now perceived that such phrases as “self government”, and the “power of the people over themselves”, do not express the true state of the case. The “people” who exercise the power are not always the same people with those over whom it is exercised.’ JS Mill: On Liberty and Other Writings in S Collini (ed) (1989).
our jurisdiction in setting standards for ‘democratic politics’ and remedying ‘political process dysfunction’?

My aim in this article is to adumbrate a theory, grounded in the idea of democratic self-government,26 that will help us better understand the Court’s ‘political process jurisprudence’. When ‘self-dealing’ representatives are corrupt or fail to act against corruption, what is really at stake, I argue, is the health of democracy as a system of self-government and democratic representation. In such circumstances, ‘democracy-reinforcing review’ overriding the usual ‘separation of powers’27 constraints imposed on the judiciary when it comes to ‘interference in the internal functioning’ of constitutionally autonomous representative bodies, may be justified by precautionary considerations, in order to vindicate the constitutional promise of self-government.

In part II, I show why the constitutional principle of ‘self-government’28 and the democratic process of collective decision-making and representation it envisages, requires ‘rule of law’ constraints which are protective of the integrity of the democratic process. This is the familiar paradox of democratic constitutionalism that Jon Elster and others have drawn attention to.29 The constitutional principle of the ‘rule of law’ is a broader notion than its administrative law derivation which is concerned with arbitrary, unfair and unauthorised exercises of discretionary powers by officials. It includes for instance a requirement that the same rules are applied to governors and the governed and that representatives act on behalf of the people and not in their own interests. In the way that I will use the concept of the rule of law, it includes all legal constraints that a self-governing community imposes over time and through experience to safeguard the integrity of the institutions and processes of political contestation, collective decision-making and representation in a self-governing political community.

In part III, I consider the question: under what legal constraints is it appropriate for the judiciary, having regard to its institutional competencies and constitutional role, to impose on the political process, designed to give expression to the constitutional principle of self-government? Answers to this question will depend on how the process of ‘will formation’

26 F Cachalia ‘Separation of Powers, Active Liberty and the Allocation of Public Resources: The E-Tolling Case’ (2005) 132 South African Law Journal 285 (Contends that the idea of self-government should be considered to be the core component of a Constitution – even one with a justiciable Bill of Rights). But see DM Davis ‘Separation of Powers: Democracy or Juristocracy’ (2016) 133 South African Law Journal 258 (Davis J responded that such recognition would encourage ‘undue deference’ in cases of rights adjudication – an odd claim since I considered it uncontentious that the principle of self-government is the primary basis for political ordering in a constitutional democracy. ‘Rights’ should not be counter-posed to the principle of ‘self-government’. Both are instantiated in the constitutions of modern democracies and are interdependent.)

27 A Vermeule Law’s Abnegation: From Law’s Empire to the Administrative State (2016) 56. Vermeule shows how the constitutional doctrine of the separation of powers has come to be understood by judges in the context of the administrative state not as a sacred constraint that must always be strictly obeyed, but rather as a principle that must be optimised. See also E Carolan The New Separation of Powers: A Theory for the Modern State (2009); William B Gwyn ‘The Indeterminacy of the Separation of Powers in the Age of the Framers’ (1989) 30 William and Mary Law Review 263. The version of the principle adopted in South African Constitutional Law is ‘flexible’, whose application case by case depends on a variety of considerations. In re: Certification of the Constitution (note 4 above) at 111. But I argue here that there must some ‘bright lines’. The normative and institutional considerations that explain each decision, where this principle is implicated, must be consistently articulated, case-by-case. The principle cannot reflect an endlessly malleable, entirely fact driven approach.

28 L Tribe The Invisible Constitution (2008) 83–91 (Offers an argument, that even if not textually explicit, both the rule of law and the principle of ‘democratic accountability’ are a part of the US Constitution).

29 Elster (note 17 above).
is understood under our Constitution, and when the political process can be considered to be constitutionally defective. While ‘process theory’ proved inadequate as a supposedly neutral account of judicial view, this article contends that it can guide judicial intervention (or where appropriate, non-intervention) in the democratic political process to remedy ‘political process dysfunction’ when it is given sufficient substantive content, grounded in the idea of representative government.

In part IV, I discuss the Court’s jurisprudence concerning the relationship between courts and representative bodies in terms of the separation of powers doctrine and respect for democratic processes. A trilogy of the most recent cases\(^{30}\) will demonstrate that the principle of the separation of powers should be invoked to preclude judicial intervention in the internal functioning of representative bodies, unless a clear violation of a specific textual provision, a constitutional obligation or a constitutional principle can be shown.

In part V, I consider what difference ‘principal/agent dysfunction’ arising from political corruption makes to the prior analysis. This kind of systemic dysfunction means that the political system is not functioning properly as a system of representative government. Judicial intervention, which displaces or qualifies the usual separation of powers considerations, may then be justified in order to vindicate the constitutional commitment to democratic self-government. The Court’s recent jurisprudence can be understood as an attempt by the judiciary to respond to the harms represented by political corruption and state capture to the constitutional project of democratic self-government. The Court has increasingly incorporated a constitutional requirement of principal/agent probity (ie, in the relationship between representatives and citizens) in its analysis of the standards the Constitution sets for the proper functioning of and judicial intervention in the democratic political process.

Although anti-corruption concerns are pervasive in these recent cases, the Court did not explicitly invoke ‘anti-corruption’ as an enforceable constitutional norm. In part VI, I conclude that an implicit anti-corruption constitutional norm should become a ‘free floating’ precautionary principle that works together with the Court’s ‘political process jurisprudence’ by clarifying when intervention is justified to remedy dysfunction in the democratic political process.

The aim of the precautionary principles in constitutional law is to protect democracy. The anti-corruption principle thus ‘has an insistently democratic foundation. Its goal is to assert popular control over the risks that concern the public’.\(^{31}\) Its adoption would justify intervention in the internal processes of the legislature in only the direst of situations – where the default assumption that the legislature is functioning as required by the Constitution cannot be made.

---


\(^{31}\) CR Sunstein The Cost-Benefit Revolution (2018) 144–155. Sunstein discusses the relationship between precautions and democracy. The subject of his book however is the regulatory state, not constitutional and democratic theory. But the precautionary principle is also important in constitutional law. It is designed to mitigate actual harms and also harms that cannot always be predicted, but if actualised, have devastating consequences for a constitutional democracy.
II SELF-GOVERNMENT, POLITICS AND THE RULE OF LAW

In her essay, ‘The Great Tradition’, Hannah Arendt says definitions of types of government always rest on two conceptual pillars: law and power. The first pillar is Platonic: It recognises that law both constitutes and constrains the exercise of power. It is recognisable in contemporary constitutionalism as the idea of the ‘rule of law, not men’: ‘lawful government was good and lawless bad’. She then contends that:

The criterion of law … as a yardstick of good or bad government, was very early replaced … by the altogether different notion … with the result that bad government became the exercise of power in the interests of the rulers, and good government the use of power in the interests of the ruled on whose behalf power is exercised. [I]t is necessary that all share equally in ruling and being ruled.

In Arendt’s description of Aristotelian politics we have an early expression of the idea of rulership in a democracy. In contemporary times, this conception of democratic rule has come to mean ‘free and equal participation in political will formation’, or government by and on behalf of the people. Modern conceptions of democratic constitutionalism combine these two ideas: (a) self-government, or ‘government by the people’ if not directly, then least by their consent and on their behalf; and (b) the ‘rule of law, not men’ – which appears at first blush to contradict the first requirement, but is actually a sine qua non for its existence.

The democratic value of self-government requires that the people are self-governing and sovereign. As Frank Michelman has observed:

Constitutionally speaking, democracy in our times … names a standard by which a country is not free, its inhabitants not free men and women, unless political arrangements are such as to place the people under their own joint rule. ‘Self-government’ it is often called.

For the people to be ‘self-governing’, they must, in some sense, be the authors of the laws to which they are subject. It guarantees rights of representation and participation in the institutions that are set up for collective decision-making.

Actualisation of the political good of ‘self-rule’ also requires a system of rules, standards, and procedures, – the legal standards encapsulated in the notion of the ‘rule of law’. The rule

33 Ibid.
34 Ibid at 58.
36 Constitution Preamble and s 43(2). Section 42(3) makes it clear that self-government in South Africa is to take the form of a representative democracy. It reads: ‘The National Assembly is elected to represent the people and to ensure government by the people under the Constitution. It does this by choosing the President, by providing a national forum for public consideration of issues, by passing legislation and by scrutinizing and overseeing executive action.’
39 J Tulley ‘The Unfreedom of the Moderns in Comparison to Their Ideals of Constitutional Democracy’ (2002) 65 Modern Law Review 204, 205 (James Tulley argues that a constitutional democracy is legitimate if it meets the test of two inextricably connected principles of legitimacy: the principle of democracy or popular sovereignty and constitutionalism or of the rule of law).
of law thus makes representative democracy possible. Self-rule or ‘active liberty’ requires a system of legally binding constraints on the political process of collective decision-making. As Montesquieu noted, a constitution (among other documents) provides a legal ‘framework within which people move and act’. This realm – the democratic political process – encompasses the representative institutions which structure collective action and party competition. The ongoing influence of this idea locally is evinced in Woolman’s recent suggestion that the Constitution is best read as a framework or scaffolding for collective action.

Hans Lindahl also sees law as a species of collective action. His ‘first person plural’ perspective on constitutions takes its departure point from the opening words, ‘We The People’, which provides:

[the] master rule that structures how a legal collective goes about responding authoritatively to the practical question: ‘What is joint action for?’ thereby determining what counts as legal behaviour for the collective.

Lindahl captures this relationship more precisely when he says:

There is a form of joint action in which the monitoring and enforcement of collective action is entrusted to certain officials and authorities. By the monitoring of joint action, I mean that in the course of joint action, certain authorities establish in a binding fashion what is its point, and how it can be best achieved, whether in the light of the changing context in which the joint action unfolds or because conflict about these issues may arise between participant agents. In such circumstances, it is up to the authorities to articulate the point of joint action by establishing what will count in the default setting of collective action, and to establish whether an act counts as a participant act.

Lindahl’s Authoritative Collective Action Model (ACA) makes it clear that harms to democratic processes of collective decision-making through which joint action and its point is decided upon are distinct from harms to individual interests. But such harms are also constitutionally cognisable. In South Africa’s constitutional democracy, ultimate responsibility for deciding on the legal criteria by which all public actions are judged falls to the Court and other independent bodies – the independent institutions created in Chapter 9 of the Constitution.

The line drawn between political accountability through the electoral process and legal accountability enforceable at a matter of law, or between ‘politics and law’, is a contextual matter. Different jurisdictions offer different answers. And the same jurisdiction might provide different answers at different times, since risk assessments to the project of democratic self-government can change over time.

Rule of law constraints are required to enable and to protect a project of democratic self-government which envisages a democratic process of representation, political contestation

---

41 Arendt (note 32 above) at 53.
43 H Lindahl ‘Constituent Power and the Constitution.’ D Dyzenhaus and M Thorburn (eds) Philosophical Foundations of Constitutional Law 145
44 Ibid at 144 (my emphasis).
45 How a constitutional system actually functions depends both on the conceptions of political justice instantiated in a constitutional text and on the traditions and practices of a particular political community. The cultural preconditions for democracy and the rule of law have to be consciously constructed and then entrenched, and re-entrenched as social mores and practices over time: they cannot simply presupposed. See J Habermas Between Facts and Norms (1998); J Rawls Justice as Fairness: A Restatement (2001) 145, 146.
and collective decision-making. The C, and lower courts and independent institutions, must intervene in the democratic political process when defects such as widespread and deeply systemic political corruption undermines the processes of genuine collective decision-making and representative government. It does so by employing the basic and founding commitment to rule of law found in s 1(c) of the Constitution. In so doing, it imposes constraints on self-dealing representatives and their private benefactors on behalf of the people and their right to democracy, as enshrined in s 1(d) of the Constitution.

However, in imposing constraints on the political process for constitutionally cognisable reasons, the judiciary is itself constrained by the constitutional principles of self-government and the separation of powers. The judiciary, as much as the legislature, is subject to the principle of constitutional supremacy set out in s 1(c) of the Constitution. The scope for judicial intervention in the democratic process to correct ‘democratic process dysfunction’ is, therefore, not unlimited.46

III JUDICIAL REVIEW AND THE DEMOCRATIC PROCESS

Judicial intervention in the internal functioning of representative bodies usually falls within some conception of the ‘separation of powers’ principle. How should the line between constitutionally prescribed intervention and usurpation be drawn? If we can identify when the political process is ‘dysfunctional’, then we can justify appropriate judicial intervention and judicial overreach. Here then is one way of articulating a norm that enacts such a distinction: the judiciary should ensure that the control exercised over other powers legitimated by the people, especially the legislature, can itself be regarded as emanating from the people and their sovereign right to self-rule.

Judicial intervention is not always, as I have pointed out elsewhere, ‘a legitimate or efficacious response to perceptions of political dysfunction’.47 In a democracy, judicial restraint in order to facilitate self-government and to show proper respect for the people’s elected representatives, will often be the right posture where no genuine issue of rights protection or rule of law observance has arisen. That said, the judiciary must intervene whenever the Constitution clearly sets out specific obligations and duties that the legislature or the executive has failed to exercise properly.

Answers to this question of line-drawing between justifiable intervention and unjustifiable overreach can also turn on how we determine when the democratic process is ‘dysfunctional’. Below, I assess the relative merits of two possible answers: (a) John Ely’s ‘representation-reinforcement’ theory; and (b) Samuel Issacharoff and Richard Pildes’ ‘structural process theory’. This assessment is not an either/or. Both theories aid our understanding of ‘democratic process dysfunction’ and justifiable judicial intervention.

A Ely’s theory of representation-reinforcing review

Ely proposed that courts should be cast in the role of guardians of the democratic process itself instead of being placed in the invidious position of second-guessing the decisions of the

---

46 EFF I (note 30 above) at para 92.
people’s elected representatives on important moral questions they should be trusted to make.  

His concern with finding a ‘democracy-reinforcing’ justification for the judicial enforcement of rights is not of much interest in South Africa. In our Constitution, the principles of constitutional supremacy and judicial enforcement of rights are clearly established in the text.

However, Ely did demonstrate that the proper functioning of the democratic process might require judicial intervention. Herein lies the contemporary relevance of Ely’s theory of ‘democratic process dysfunction’ for non-rights cases in South Africa. His insights enable us to draw a connection between the idea of popular self-rule and the case for ‘representation-reinforcing review’ to remedy ‘democratic process dysfunction’.

Ely built his theory on the back of Carolene Products, which identified ways in which the ordinary democratic process in a system of representative government can malfunction. Ely assumed that the representative political process envisaged by the US Constitution was ordinarily majoritarian and based on efficient pluralist bargaining. But where these assumptions break down and the process can therefore be said to be defective, the judiciary can justifiably intervene in order to protect those effectively unable to protect themselves through pluralist bargaining in a representative, democratic process.

His theory of representation-reinforcement also emphasises principal/agent accountability between the people and their representatives. This second theme recognises that the relationship between representatives and citizens/voters can also break down in practice when representatives act in ways which are self-serving and constitute an abuse of democratically legitimate incumbency. So, in addition to the judiciary’s role in protecting the rights of ‘discreet and insular minorities’, it has the responsibility of protecting the democratic character of representational institutions so as to ensure their accountability to the people.

Ely’s process theory of judicial review, so read, can provide a theoretical starting point for identifying a class of constitutionally cognisable process harms. They encompass harms that representatives (often in concert with private actors) pose to legitimate democratic decision-making. ‘Democracy-reinforcing judicial review’ applies to both harm to the democratic process and the harmful consequences for individuals and minorities that might result from a ‘dysfunctional process’. The kind of systemic political corruption – so-called capture of the democratic political process – that South Africans have experienced for the last score of years impairs the system of accountability and representation explicitly envisaged by the Constitution. In response to such conditions, the judiciary must craft constitutional remedies designed to restore, in so far as possible, the proper workings of our democracy.

---

51 Ely (note 49 above) at 77.
52 Ibid at 78.
53 See notes 25 and 37 above.
B Structural process theory and ‘self-dealing’

In the US, judicial supervision of the political process stretches back at least to the early 1960s and electoral rights. Baker v Carr served as precedent for applying the principle of one person, one vote to the abuse of incumbency via gerrymandering as well as the fairness of campaign finance restrictions. Gerrymandering and campaign finance invariably raise concerns about political corruption of the democratic process. Pildes has described this ‘largely unappreciated transformation in constitutional law’ as the ‘constitutionalization of democratic politics’. He contends that the dominant mode of constitutional theorising – which places individual rights at the centre – does not provide an adequate analytical framework for understanding the constitutional questions that are implicated when it is the functioning of the democratic process itself which requires consideration.

Pildes and Issacharoff have tried to solve that inadequacy with a theory that relies on a compelling analogy between competition in political markets and competition in corporate markets (governed by antitrust law). The theory targets abuses of dominance – not dominance per se – by ‘insiders’. It concentrates its focus on the protection of electoral processes in a manner that ensures accountability by eliminating problems of self-dealing by political incumbents designed to entrench themselves in power through illicit means.

However, my concerns regarding ‘political process dysfunction’ naturally differ from those that animate Pildes and Issacharoff’s project. They are concerned with ‘partisan lockups’ and other forms of electoral engineering by incumbents that undermine fair electoral competition. My principal/agent driven theory concerns itself predominantly with the twin ideals of ‘democratic self-government’ and ‘accountability of representatives to the people’. My approach converges with that of both Ely and Pildes and Issacharoff by emphasising that judicial review can play a ‘democracy-reinforcing role’ that cabins ‘self-dealing’ by incumbent representatives. Process theory as a distinct genre of constitutional theorising provides valuable insights into the question when the democratic political process can be considered to be functioning in accordance with constitutional dictates, and when they fall afoul of the basic law’s provisions.

IV THE CONSTITUTIONAL COURT AND THE SEPARATION OF POWERS
IN MATTERS PERTAINING TO THE PROCESSES OF REPRESENTATIVE BODIES

The South African judiciary has intervened increasingly in the internal processes and functioning of legislatures. These ‘internal matters’ take many forms, including: the exercise of disciplinary powers by Parliament over its members, its compliance with a constitutional obligation to facilitate public participation in its legislative processes, the treatment of ‘private

54 Baker v Carr 369 US 186 (1962) (In this landmark decision, the US Supreme Court held that the political question doctrine does not prevent courts from assessing the fairness of redistricting, especially when gerrymandering vitiates the basic democratic and constitutional commitment to one person, one vote).
58 Speaker of the National Assembly v De Lille 1994 (4) SA 863 (SCA).
members bills’, 60 the scheduling of motions of confidence, 61 the powers of the Speaker to require a secret ballot, 62 and, finally, the need to promulgate rules regulating impeachment of a President as well as the composition of a parliamentary committee convened to consider the impeachment of a President. 63

Cumulatively, these decisions have significant implications for the relationship between the judiciary and Parliament. They depart from widely accepted understandings of the constitutional norms and principles governing the relationship between courts and legislatures in more established constitutional democracies. Indeed, the differences are profound – whether the legislature takes a ‘Westminster’ 64 form, reflects the ‘constrained parliamentarism’ of Canada 65 and Namibia, or looks like a US style system with its broad array of ‘checks and balances’. In these different jurisdictions, the internal proceedings are generally, though not always, considered to be generally unreviewable. 66 The South African decisions by contrast constitute – as Michael Walzer might describe them – ‘deep penetration raids into the area of legislative decision’ 67 and are ‘radically intrusive on what might be called democratic space’. 68 Put differently, they curtail majoritarian decision-making on questions usually reserved for the legislature. 69

The question arises as to how these incursions can be justified. 70 As Aziz Huq notes in this volume, the Court interventions may simply be tactical attempts to safeguard the democratic process – and thus not readily susceptible to explanation by a single, overarching theory. 71

60 Oriani-Ambrosini, MP v Sisulu, MP Speaker of the National Assembly [2012] ZACC 27, 2012 (6) SA 588 (CC) (‘Ambrosini’).
61 Mazibuko v Sisulu & Another [2013] ZACC 28, 2013 (6) SA 249 (CC) (‘Mazibuko’).
62 UDM (note 30 above).
63 EFF 2 (note 30 above).
65 In Canada the legislature remained internally autonomous even after the introduction of the Charter of Rights. The willingness of the South African Constitutional Court to intervene in the internal functioning of Parliament may well be consistent with trends in jurisdictions which have more recently adopted the principle of constitutional supremacy. Mali introduced a centralised system of judicial review in 1992. The Malian Constitutional Court has intervened in the country’s legislative processes in order to strengthen minority rights and the independence of individual members from party control. See L Heeman ‘Judicial Review and the Democratization in Francophone Africa: The Case of Mali’ (2018) 51 Verfassung und Recht in Ubersee 166. Judicial review, as a technique for ensuring the fairness of ‘democratic politics’, does not have a fixed content and is clearly jurisdiction dependant.
66 Compare R Colker & J Brudney ‘Dissing Congress’ (2001) 100 Michigan Law Review 80 (Critiques the US Supreme Court for its insistence that Congress develop a record in support of proposed legislation: such insistence amounts to undue interference in the internal functioning of a co-equal branch).
68 Ibid.
69 But see Elster (footnote 17 above), who remarks: ‘Legislatures are to a large extent, although far from completely, masters of their own rules of order’ (at 21). What is remarkable in the evolution of South African constitutionalism is the extent to which the Legislatures are not ‘masters of their own rules of order’.
70 Judicial intervention may be required to contain the effects of executive dominance in parliamentary systems, see Gardbaum (note 10 above). See also U Beck Risk Society: Towards a New Modernity (1992) 188 (Ulrich Beck identifies ‘a shift of former Parliamentary powers to factions and parties’ and the ‘automization of the state apparatus against the will of citizens’).
My democracy-reinforcing theory proffers only one possible form of justification and critique. In this part of the article, I examine the Court’s reasoning from what might be characterised as a ‘traditional’ separation of powers perspective. In part V, I consider whether the democratic process can be considered to have ‘malfunctioned’ in a trilogy of cases that came before the Court over the last two years, and what difference that might make to the analysis. I suggest how ‘democracy-reinforcing judicial review’ might justify overriding ‘traditional’ separation of powers considerations when the political process can be said not to have functioned in accordance with constitutionally prescribed standards.

A The Separation of powers

In EFF 1, the Court addressed two novel issues: (a) whether the Public Protector’s remedial orders have binding legal effect; and (b) whether the National Assembly had an obligation to hold the President accountable in light of the Public Protector’s adverse findings. Given that the Court held that the Public Protector’s remedial orders have binding legal effect, it had some work to do with respect to the National Assembly’s resolution absolving the President of any adverse findings in the Public Protector’s Secure in Comfort report.72 However, given that the Public Protector’s remedies and the National Assembly’s resolution were not technically incompatible with one another, the Court found itself able to give full effect to the separation of powers principle. It held that it fell ‘outside of the parameters of judicial authority to prescribe to the National Assembly how to scrutinise executive action, what mechanisms to establish, and which mandate to give them, for the purpose of holding the executive accountable.’73

The Court soon found itself confronted with a similar separation of powers issue in EFF 2. In this matter, however, the Court did ‘prescribe to the National Assembly how to scrutinize executive action’ in order to fulfil its institutional function ‘of holding the executive accountable’. It delineated the requirements with which impeachment rules must comply: (a) the establishment of a ‘specific mechanism’ in the form of a standing committee; (b) the creation of a permanent committee as opposed to ad hoc committee; and (c) the composition and the mode of decision-making of this ‘mechanism’, so as to ensure that the committee was not controlled by the parliamentary majority.74

---

72 Secure in Comfort (note 20 above).
73 EFF 1 (note 30 above) at para 93.
74 Adrian Vermeule ‘Submajority Rules: Forcing Accountability upon Majorities’ (2005) 13 Journal of Political Philosophy 74. Adrian Vermeule contends that sub-majority rules on procedural questions can, and should, be employed in order to enable minorities to hold majorities accountable and to ensure that their decision-making is transparent. This argument provides support for the Court’s reasoning in EFF 2, as well as other decisions such as Ambrosini and Mazibuko. Justice Froneman offers an alternative account of what the Court had done. In his concurrence, Justice Froneman argues that the Court had done no more than provide ‘guidance’ to the National Assembly as to how it should discharge its constitutional obligations. EFF 2 (note 30 above) at para 284. He writes: ‘It attempts to provide the National Assembly with the tools necessary to enable it to fulfil its constitutional duty, to hold the President accountable in the direct of situations. It does not tell the National Assembly how to use those tools.’ EFF 2 (note 30 above) at 285 (my emphasis).
As a textual matter, s 89 of the Constitution prescribes no specific rules, procedures or mechanisms for impeachment. It grants the National Assembly the authority to impeach the President and sets out three grounds for discharging the President from office. But it nowhere determines the procedures demanded to carry out this exceptional exercise of political power.

The Chief Justice, in dissent, complained that the majority’s holding had strayed into the ‘exclusive domain of Parliament’ and that its order constituted ‘a textbook case of judicial overreach’. As for textbook cases, in *Ambrosini* the Chief Justice held not only that Parliament was constitutionally obliged to adopt rules regulating the treatment of ‘private members bills’ but that individual members had a constitutional right to have such legislation drafted at Parliament’s expense, and to compel consideration of such legislation, even if there were no prospects of gaining the support of the majority. In *Mazibuko*, the Court held that Parliament was obliged to adopt rules regulating the scheduling of motions of confidence, and preventing the majority from delaying consideration of such motions – and the Chief Justice concurred. In neither *Ambrosini* nor *Mazibuko* did the Court rely on any specific direction in the text of the Constitution (or even a well established constitutional norm). In all three cases, the Court’s actions reflected clear ‘intrusions’ into the internal rule-making authority of the National Assembly. Section 57(1) of the Constitution states:

> The National Assembly may
> (a) determine and control its internal arrangements, proceedings and procedures;
> (b) make rules and orders concerning its business, with due regard to representative and participatory democracy, accountability, transparency and public involvement. which is expressly protected by the Constitution.

_Ambrosini, Mazibuko_ and _EFF 2_ are all arguably ‘textbook cases of judicial overreach’ on ‘traditional’ separation of power grounds.

The _EFF 2_ Court’s order requiring the Assembly to draft rules for impeachment did not, ostensibly, dictate ‘content’. However, the Court held that a standing committee – as distinct from an ad hoc committee – on impeachment was to be established. In addition, the reasoning in the main judgment made it clear that such a committee would have to be controlled by the minority in order to be constitutionally compliant. Again, there is little textual basis for an outcome that clearly encroaches upon the autonomy of the representative branch. In conclusion, _EFF 1, EFF 2, Ambrosini_ and _Mazibuko_ demonstrate that although the Court remains concerned with separation of powers considerations, it has yet to develop a normative framework that accounts coherently for those decisions that respect, and those that override,

---

75 Constitution s 89 reads, in relevant part, as follows: Removal of President 89. (1) The National Assembly, by a resolution adopted with a supporting vote of at least two thirds of its members, may remove the President from office only on the grounds of— (a) a serious violation of the Constitution or the law; (b) serious misconduct; or (c) inability to perform the functions of office.

76 _EFF 2_ (note 30 above) at para 223.

77 *Ambrosini* (note 60 above).

78 *Mazibuko* (note 61 above).

79 Cachalia (note 47 above).

80 _EFF 2_ (note 30 above) at paras 191 and 192.
the ‘rule-making’ powers exercised by the National Assembly and when it regulates its own ‘internal-functioning’.  

B The political process and the value of political equality

Democratic electoral processes are, uncontroversially, based on the egalitarian principle of ‘one person, one vote’. But what principle of equality should govern collective decision-making in representative bodies where all law and all state conduct draws its force from the Constitution and no law nor state conduct is immune from judicial review?

Let’s step back for one moment. Equality in a democracy takes three forms.

The first, political equality, has four facets. Each adult citizen exercises the franchise. Every adult citizen – unless she has run afoul of the law in particular ways – may run for office. Each person possesses one vote, and no more. Governments in a parliamentary democracy are elected in a two-stage process: by a majority of votes cast in an election and subsequently by those elected representatives who constitute the legislative branch, ie, by a ‘bare majority’.

Political equality however also has two egalitarian background conditions. Legal equality, as s 9 of the Constitution tells us, requires that all persons are equal before the law. In short, no one may face discrimination on ascriptive characteristics such as race, ethnicity, sex, religion, age, birth, or disability. Moral equality requires that each person in a polity must be treated (a) by the state with equal concern and respect; and (b) with dignity by his or her compatriots.

Most representative bodies take legislative and non-legislative decisions, by majority vote. In most constitutional democracies – and South Africa is no different – minority parties,

81 The problem of ‘inadequate procedures’ has plagued impeachment trials in the US. R Posner An Affair of State: The Investigation, Impeachment, and Trial of President Clinton (1999) 120. The rules, procedures and practices related to impeachment in the US remain unsettled despite the existence of a 230-year Constitution. Perhaps their absence can be attributed to only two successful attempts to impeach a President in the House, both of which failed in the Senate. L Tribe & J Matz To End A Presidency: The Power of Impeachment (2018). Laurence Tribe and Joshua Matz’s work offers a cautionary tale regarding impeachment proceedings against an elected President: ‘[E]nding the Presidency … is a very big deal … [W]e have no historical experience with the full consequences of pushing the red button’. Despite Tribe’s initial reluctance to pursue this potentially explosive path, after the publication of the Mueller Report, he revised this position in an op-ed piece published in USA Today (21 April 2019). See also L Greenhouse ‘The Impeachment Question’ LXVI New York Review of Books (June 27 2019) 22.

82 Fedsure Life Assurance Ltd v Greater Johannesburg Transitional Metropolitan Council 1999 (1) SA 374 (CC), 1998 (12) BCLR 1458 (CC) at paras 58–99 (‘It seems central to the conception of our constitutional order that the legislature and executive in every sphere are constrained by the principle that they may exercise no power and perform no function beyond that conferred upon them by law.’) Ex Parte President of the Republic of South Africa: In re Pharmaceutical Manufacturers Association of South Africa (2000) (2) SA 674 (CC), 2000 (3) BCLR 241 (CC) at para 33 (‘The control of public power by the courts through judicial review is and always has been a constitutional matter. Prior to the adoption of the Interim Constitution this control was exercised by the courts through the application of common law constitutional principles.’)

83 The proportional ‘closed list’ electoral system adopted in South Africa respects this principle and is not susceptible to the ‘vote dilution’ that can occur through the manipulation of electoral boundaries in some constituency-based systems. Nor is any vote ‘wasted’.

84 Constitution ss 9 [equality] and 19 [political rights].

85 Ibid ss 42, 46, 47 and 86.

86 Woolman (note 37 above) at 374.

87 Where this is not the case, as for example for constitutional amendments and impeachment, the Constitution is specific.
minority coalitions or transient minorities still participate in the process of decision-making as of right. 88 Consistent with the above account, Jeremy Waldron argues that majoritarian decision-making in legislatures is not merely a technical device, but rather a fundamental egalitarian principle of political morality which treats each participant as an equal. 89 The Constitution also provides emphatically that all questions before the National Assembly are to be decided by a majority of votes cast. 90

We now have a working definition of political equality which has a constitutional basis. With it we can take accurate measure of the most recent and troubling case law.

In EFF 1 the Court decided unanimously that a resolution passed by the National Assembly absolving the President of compliance with the remedial action taken by the Public Protector in the ‘Nkandla’ matter was inconsistent with the Constitution. As far as this author is aware, the judgment constitutes the first occasion that a procedurally correct motion passed by a majority of votes cast in the National Assembly has been found to be unconstitutional. It is a curious conclusion given that s 53(1)(c) read with s 57(1)(a) of the Constitution suggests that all questions before the National Assembly can be decided by a majority of the votes cast, except where the Constitution specifically provides otherwise. This article revisits this question in part V.

In EFF 2, the Court held that the National Assembly was in breach of its constitutional obligations to initiate impeachment proceedings and to draft rules for the impeachment process in terms of s 89 against the President. Again, the Court so held despite the fact that the President enjoyed support of a political majority and that the Constitution leaves such decision-making to the National Assembly by majority vote. Impeachment is the most serious and exceptional instrument of parliamentary censure and is fraught with risk to the integrity of a representative democracy since it potentially results in the displacement of an elected President through an invariably partisan, party political process. Not surprisingly, the nature of the risk generated strong disagreements within the Court about whether it was appropriate for the judiciary to insert itself in the process and to countermand the decisions of the parliamentary majority.

Parliament had done nothing to delay or to impede the tabling or consideration of several no confidence motions and a motion to impeach. As a result, no need arose to create rules that specifically regulates impeachment. When such a question arose in the past, an ad hoc

88 Section 57(2)(b) allows minority parties to participate in the proceedings of the National Assembly ‘in a manner consistent with democracy’. Democratic Alliance v Masondo CCT 29/02 [2002], ZACC 287 (2003)(2) BCLR 128, 2003(2) SA 413(CC)(Court held that minority parties in a Municipal Council were not entitled to representation on the Mayoral Committee under the provisions of the Municipal Structures Act 117 of 1998.)

89 J Waldron Democracy and Disagreement 107. Compare K Arrow Social Choice and Individual Values (1963). Arrow’s ‘impossibility theorem’ shows why no decisional apparatus, especially one based upon majority rule over non-binary policy choices, reflects accurately the aggregation of all individual preferences. However, some systems diminish this problem through a combination of proportional representation and party discipline. K May ‘A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decision’ (1952) 20 Econometrica 680. Of course, the problem of choice is replicated within political parties: members will disagree over an array of policies. See also CR Beitz Political Equality (1989). Beitz argues that majority rule ‘is wrongly conceived as expressing a foundational requirement of political equality. Its significance is far more modest’ (at xii). But he does not apply his conception of political equality and ‘complex proceduralism’ to decision-making in legislative institutions.

90 Constitution s 53(1)(c).
committee was established on the initiative of the parliamentary opposition. The absence of specific parliamentary rules on impeachment and provision for the establishment of a standing committee to handle impeachment proceedings could hardly constitute an impediment to an impeachment inquiry when no majority felt the need to proceed with such a momentous interrogation of the President. Parliaments across most jurisdictions do not allocate their scarce resources—time, staff and money—on speculative endeavours. That is one obvious reason that ostensible ‘gaps’ in the rules of Parliament always exist. Political adversaries will also defer decisions on contentious issues. For that reason, parliamentary rules are always ‘incompletely theorized’.

Prior to EFF 2, the Rules Committee had in fact set up a subcommittee to draft impeachment rules but were never adopted. Opposition parties failed to make submissions. Why? The committee was controlled by the majority party, as are all parliamentary committees. They therefore considered litigation to provide better prospects of success. However, the principle of majoritarian decision-making on ‘internal matters’ has been constitutionalised in the Constitution. Thus, the fact of majoritarian control of a parliamentary committee cannot alone be a source of constitutional concern. In EFF 2, a majority of the members of the President’s party did not agree with the opposition—politically—that the President should be impeached. That is representative democracy. Absent a two-thirds majority, the motion to impeach the President, tabled by the opposition, failed. Not surprisingly, Deputy Chief Justice Zondo’s dissent repeatedly asked for a constitutional basis to compel the National Assembly to initiate impeachment proceedings against a President who retained the political support necessary to remain in office. Likewise, the Chief Justice’s opinion turned on a separation of powers argument and understandings of what political equality in a representative democracy requires.

The majority’s justifications for its circumvention of majority-rule with respect to decision-making procedures appear somewhat opaque. Justice Froneman writes cryptically that all the Court had done was to require the National Assembly to ‘hold the President to account in the direst of situations’. He left unexplained those considerations that had led him to reach this assessment. Instead, the Court would ‘leav[e] it to history to determine whether the order in the second judgement … achieve[d] its aims’. The EFF 2 Court does no more than contend that the National Assembly was in breach of its obligations to hold the President accountable in terms of s 89 of the Constitution given the findings in EFF 1 that the President had violated the Constitution by disregarding the Public Protector’s legally binding remedies.

---

91 The committee was created in response to the government’s decision to allow President Omar al Bashir, President of Sudan to leave South Africa. Under the Rome Statute, al Bashir, who had been indicted by the International Criminal Court, should have been handed over to the Tribunal.


93 Justice Jafta’s argument against the sufficiency of an ad hoc committee is that such committees are controlled by the majority. Actually, all committees, whether ad hoc or standing, are generally controlled by the majority. The Constitution itself confers powers on the National Assembly, in terms of s 57 to make rules and orders regulating the duration and composition of committees, and requiring only that provision be made for participation in such committees by minority parties ‘in a manner consistent with democracy’. This begs the question. Had the National Assembly adopted rules to regulate the impeachment process which provided for the establishment of a standing committee controlled by the majority, and making provision for minority participation consistent with electoral outcomes, would such rules in the view of Justice Jafta have been constitutionally deficient? It is difficult to reconcile such a conclusion with the with specific textual provisions.

94 EFF 2 (note 30 above) at para 285.

95 Ibid at para 286.
However, it does not follow – logically, textually or politically – that the outcome in EFF 1 automatically triggered a constitutional obligation to create all the procedural trapping required for impeachment proceedings against the President in the Assembly, and to initiate such proceedings. Nor can the Court claim that EFF 2 simply reiterates the reasoning or the holding in EFF 1. The EFF 1 Court offers no prescription and no signal that its decision contemplates such a trajectory. Indeed, that the unanimous decision in EFF 1 is followed by a 5–4 split in EFF 2 strongly suggests that EFF 2 does not follow the core logic and concerns of EFF 1.

EFF 2 reflects, at least implicitly, different conceptions of the separation of powers doctrine and of representative democracy. Justice Zondo writes:

It’s not a bad thing for different members of the National Assembly to hold different views on any issue. Our constitution expects there to be different views on issues in the National Assembly. This is why it provides in section 53(1)(c) that all questions before the Assembly are decided by a majority of the votes cast.96

Whilst maintaining that the National Assembly had, to some degree, held the President accountable,97 his opinion is primarily animated by a concern that the Court’s order ‘bypasses the democratic process’ envisaged by the Constitution. By allowing minority parties to use the courts to achieve ends that they could not achieve in the National Assembly, the Court ignores unambiguous textual requirements that all decisions by the National Assembly related to the establishment and composition of committees must be initiated in the Assembly by majority vote.98 To be fair, a reading exists that makes some sense of the majority judgment. As in Modderklip,99 and EFF 1, the majority of the Court appear concerned that such a fundamental abrogation of the rule of law had occurred that it raised questions as to whether the ‘basic structures’ of the country’s democratic institutions remained intact.100 If so concerned, the majority had an obligation to state this basis for their counter-majoritarian intrusion crystalline clear.

Part of the answer to the question as to whether and when counter-majoritarian constraints on ‘non-legislative’ parliamentary decision-making can be justified, is that the Constitution also

---

96 Ibid at para 62.
97 Ibid at para 89.
98 Constitution s 53(1)(c)(note 90 above).
99 President of the Republic of South Africa & Another v Modderklip Boerdery (Pty) Ltd [2005] ZACC 5, 2005 (5) SA3 (CC) (‘Modderklip’).
100 United Democratic Movement v President of the Republic of South Africa [2002] ZACC 21, 2003 (1) SA495 (CC).
contains competing normative commitments: to minority rights  and freedom of speech. This trilogy of cases raised neither such concern. Again the majority only hints at another constitutional rationale that arises in the ‘direst of situations’.

In part V, this article looks at this trilogy of cases – EFF 1, UDM, and EFF 2 – through the lens of the principal/agent relationship envisaged by the Constitution between representatives and voters, the principle of accountability implicit in any conception of representative democracy, and expressly set out in the Constitution. In this light, the article explores the implications of the failure of parliamentary processes over a protracted period, to hold the President to account for the misuse of public funds and for his orchestration of ‘state capture’. This kind of systemic ‘democratic process dysfunction’ provides occasion for ‘democracy-reinforcing review’ in order to reinforce the accountability of representatives to the people who elected them. However, the case made in part V is highly qualified. Parliamentary or legislative autonomy, in a constitutional system committed to the separation of powers, should generally assume that majoritarian parliamentary processes are ‘functioning well’.

---

101 Constitution s 57(2) states: ‘The rules orders of the National Assembly must provide for (a) the establishment, composition, powers functions, procedures and duration of its committees; and (b) the participation in the proceedings of the Assembly and its committees on the minority parties represented in the Assembly, in a manner consistent with democracy.’

102 Constitution s 58. Free speech in Parliament poses some interesting conundrums in our jurisdiction. What should happen if a Speaker, who has not been properly briefed on the requirements of s 58(1)(a) and (b), seeks to regulate the constitutionally protected content of speech by the government’s parliamentary opponents? On my democracy-reinforcing process theory, it would reflect a ‘clear case of dysfunction’. Freedom of speech in Parliament is a fundamental constitutional value, which must be observed in the practices of the National Assembly, when it is functioning as envisaged by the Constitution. Where this clear constitutional commitment is not observed, the National Assembly cannot be said to be functioning in a way that is constitutionally compliant. When the Court intervenes in such circumstances, it relies on a specific provision of the Constitution – parliamentary freedom of speech. However, absent such a clear constitutional basis for intrusion, it should be recognised that judicial supervision of parliamentary proceedings can have ‘disabling effects’ on the functioning of representative bodies. New Brunswick Broadcasting Co. v Nova Scotia (Speaker of the House of Assembly) [1993] 11 SCR 319 (Canada). In New Brunswick, Justice McLachlin explains that the exercise of powers to review decisions of the Speaker would impair the ‘dignity and efficiency’ of the legislature: ‘Quite apart from the constitutional question of what right the courts have to interfere in the internal process of another branch of government ... The ruling of the Assembly would not be final. The Assembly would find itself caught up in legal proceedings and appeals about what is disruptive and not disruptive. This in itself might impair the proper functioning of the Chamber.’ Ibid at 383.

103 Section 1(d) read together with the provisions regulating the structures of representation: ss 42(3), 46, 47 and 55.

104 The question whether a representative body is ‘functioning well’ will often be in the eyes of the judicial beholder. In assessing the functioning of the political process, judges should recognise that parliamentarians, whether in the majority or minority, act, or fail to act, for many different reasons which are beyond judicial consideration or scrutiny, especially on questions of motions of confidence or impeachment. Their ‘political judgments’ should therefore be accorded a wide ‘margin of appreciation’. The often messy business of politics, characterised by conflict and disagreement, must be allowed the space to ‘work’. At the same time, judicially cognisable instances of systemic dysfunction in the structures of democratic representation most certainly exist. The Constitution, I argue below, sets at least some judicially enforceable standards on the conduct of representatives in order to reinforce what I have called principal/agent accountability between elections.
V PRINCIPAL/AGENT ACCOUNTABILITY AND DEMOCRATIC PROCESS DYSFUNCTION

A Principal/agent accountability

On 19 March 2014, the Public Protector issued her report into allegations of impropriety and unethical conduct as they related to the significant expenditure of public funds on the President’s private homestead at Nkandla. The report’s rigorous analysis of the evidence buttressed its damning findings of non-compliance by the President with Cabinet policy, ethics requirements, and public procurement legislation. It held that a broad array of senior officials and Cabinet Ministers had engaged in related and similar illegal and irregular behaviour.

The findings against the President constituted a serious indictment of his conduct: (1) he and his family were found to have ‘unduly benefited from the enormous capital investment from non-security installations at his private residence’; (2) he had ‘failed to discharge his responsibilities’ as the ‘guardian of the resources of the people’; (3) he was in breach of his ethical obligations under the Executive Members Act and Code; and (4) his conduct, especially with regard to protecting public resources, was ‘inconsistent with his office as a member of Cabinet, as contemplated by s 96 of the Constitution’. The constitutional basis for the final finding, s 96, provides that members should not ‘act in any way that is inconsistent with their office’, expose themselves to any situation involving the risk ‘of a conflict between their official responsibilities and private interests’ or use their position ‘to enrich themselves or improperly benefit any other person’. The President was ordered to repay a reasonable percentage of the cost, to be determined by the National Treasury, to reprimand the Ministers involved, and to report to the National Assembly within 14 days with regard to his illegal and irregular ‘actions [found in] this report’.

The President should have implemented the remedial action or sought to review the Public Protector’s remedial action in court if he disagreed with the findings. Instead he directed the Minister of Defence, in a report to the Speaker, to review various policies relating to presidential security and to determine ‘whether the President is liable for any contribution in respect of security upgrades’. This investigation and the reports of two parliamentary committees controlled by the President’s party predictably absolved the President from all wrongdoing and responsibility. The National Assembly then passed a resolution by majority vote ‘effectively nullifying the findings made and remedial action taken by the Public Protector’. By passing that resolution the National Assembly had ‘flouted its obligations’.

This factual matrix plays at critical role in understanding the context in which the EFF Court came to the conclusion – almost two years after publication of the report and thus two years of non-compliance – that the Public Protector’s remedial orders had binding legal effect. While Woolman notes that this result can be justified with reference to the ‘constitution’s

---

106 Secure in Comfort (note 20 above) at para 10.9.1.4.
107 Ibid at para 10.10.1.4.
109 Secure in Comfort (note 20 above) at para 10.10.5.
110 Ibid at para 11.
111 Ibid at para 63.2 (my emphasis).
112 Ibid at para 98.
113 Ibid at para 99.
cleverly structured design’114 he makes a far more incisive set of observations about how the law and the factual substratum interact. He writes:

As the mercenary behaviour of our state officials became increasingly visible, the ongoing absence of any efficient, anti-corruption institution up to the task of fighting such widespread corruption became evermore evident. It is therefore not surprising that our appellate courts read the basic law in a manner that recognized a binding legal effect of the Public Protectors recommendations.115

The EFF 1 Court further explained why the Public Protector’s remedial action must be binding:

[The Public Protector] is the embodiment of the Biblical David … who fights the most powerful and very well resourced Goliath – the very incarnation of impropriety and corruption by government officials. The Public Protector is one of the true crusaders and champions of anti-corruption and clean governance.116

For David to defeat Goliath, the Public Protectors powers must ‘be very wide’ – and such breadth is further justified, because in terms of the rule of law, ‘no lever of governmental power is above scrutiny and censure’.117 When it came to the plundering of the public purse and the pillaging of the public’s source of basic goods, the Court and other independent state actors (for example, Chapter 9 institutions) had constitutional obligations to put a halt to such widespread, venal behaviour because:

[the] repositories of these resources and power are to use them for the benefit of the public. When this is suspected or known not to be so, then the public deserves protection and that protection has been constitutionally entrusted to the Public Protector.118

The EFF 1 Court’s powerful statements restate how this basic relationship between citizens and their representatives in a constitutional democracy is undermined by ‘self-dealing’, the abuse of power and political corruption.119

In addition, the EFF 1 Court also held that the resolution passed by Parliament – while proper in form and by a majority vote – that absolved the President of liability had been unconstitutional. By failing to hold the President to account in circumstances where there had been compelling evidence before the Court of abuse of power and misappropriation of public resources, the National Assembly violated its constitutional obligations in terms of s 42(3) of the Constitution.120 This section reads as follows: ‘[The] National Assembly is elected to represent the People and to ensure government by the people under the Constitution’. The Court rightly reasoned that the National Assembly carries out this function ‘by choosing the President … and by scrutinizing and overseeing executive action’.121 In reaching this conclusion, the Court

115 Ibid at 166 (My emphasis – ‘appellate courts’ refer to the Supreme Court of Appeal (in SABC v DA) and the C).
116 EFF 1 (note 30 above) at para 52.
117 Ibid at para 53.
118 Ibid at para 53.
119 The Court grounded its holding in the text of the basic law by finding that the President’s failure to comply with the remedial orders was inconsistent with s 83(b) read with ss 181(3) and 182(1)(c) of the Constitution. It ordered him to repay a reasonable percentage of the non-security costs incurred on his private home, to be determined by the National Treasury. The EFF 1 the Court explained that the President is the ‘personification of the nation’s constitutional project’ and that he is to ‘serve the people well and with integrity’. Ibid at 20 and 21.
120 Ibid at paras 94 and 98; and para 10 of the Court’s order.
121 Ibid at paras 97 and 104.
demonstrated that – based on the rule of law, its co-foundational value of constitutional supremacy and the separation of powers doctrine – it (a) had the responsibility to set criteria by which all public actions are judged; and (b) would enforce a constitutional standard of accountability for the misappropriation of public resources against both the executive and the parliamentary majority responsible for oversight where and when it was so required.122

This constitutional standard requires the National Assembly to act in ‘good faith’ as the representative of the people and the custodian of their resources. In this matter, it had not functioned as envisaged by the Constitution. The EFF 1 Court anticipated a separation of powers objection, given that it had countermanded a decision of the parliamentary majority and thereby ‘intruded’ upon the National Assembly’s space. The EFF 1 Court explained that all it had done was to determine if the National Assembly had acted, in ‘substance and reality’,123 in a manner consistent with the Constitution. Importantly, for this line of reasoning, the EFF 1 Court did not prescribe how and ‘by what structures or measures’124 the National Assembly should carry out its legal obligations.

In two subsequent cases, the Court continued to invoke a requirement of principal/agent accountability. In UDM, the Court had to determine whether the Speaker had the constitutional power to prescribe that voting in a motion of no confidence in the President be conducted by secret ballot.125 The motion arose out of the continuing controversy following the Public Protector’s Secure in Comfort report,126 the finding in EFF 1 that the remedial action had binding legal effect, and the ongoing failure of the National Assembly to hold the President to account. The Speaker, probably acting on the misguided advice of the table staff, had taken the view, in response to a request from the applicant, that she did not have this power either in terms of the Constitution or under the National Assembly’s rules, to order a secret ballot.

The purpose of the secret ballot would be to test whether the President still had the support of the majority of members able to vote ‘in accordance with the conscience of each’127 when implicit threats of punishment for breaches of ‘party discipline’ were removed.128 In coming to the conclusion that the Speaker did possess this power, the UDM Court began as follows:

South Africa is a constitutional democracy – a government of the people, by the people and for the people though the instrumentality of the Constitution. It is a system of governance that ‘we the people’ consciously and purposefully opted for, to create a truly free, just and united nation. … [The Constitution reflects the understanding] that it is not practical for all 50 million of us to assume governance responsibilities … and that ‘we the people’ designated messengers or servants to run our constitutional errands for the common good of us all. … [Judges too are] servants of the people.129

The UDM Court continued as follows:

[Parliamentarians are] another set of servants. … These servants are supposed to exercise the power and control these enormous resources at the beck and call of the people. Since state resources
are for our common good, checks and balances to ensure accountability enjoy pre-eminence in our governance system.¹³⁰

The servant metaphor is not simply a superfluous embellishment in the Court’s reasoning: by distinguishing good from bad power, it functions as a criterion of law,¹³¹ or in other words, the rule of law.

In UDM, the Court held that although the Constitution did not expressly authorise the Speaker to order a secret ballot in these circumstances, she retained the power to do so. In exercising this discretion, however, she was required to pay due regard to ‘real possibilities of corruption’.¹³² By pressing the Speaker so, the UDM Court reinforced principal/agent accountability as a criterion of constitutionally authorised ‘rulership’.

EFF 2 was handed down some 14 months after the Public Protector’s second major report: The State of Capture.¹³³ The document revealed the extent of systemic corruption and how it threatened the basic structures of representative democracy and constitutional government. At the time of the hearing, the National Assembly had still not taken the remedial action required by the Public Protector’s Secure in Comfort report. This raised the question whether both the National Assembly and the President were in compliance with their constitutional obligations. The Court held that the failure of the National Assembly to make specific rules regulating presidential impeachment and to initiate impeachment proceedings against him, was inconsistent with the principal/agent relationship envisaged by the Constitution, since the National Assembly is elected to represent the people, and must ensure:

[government] by the people by scrutinising and overseeing executive action. It also achieves this purpose by choosing the President and providing a national forum for public consideration of issues. This underscores the role played by the Assembly as the people’s representative.¹³⁴

It might be added: it carries out its representative mandate by impeaching an elected President who threatens the foundations of the constitutional order by setting up a corrupt system, and immunising himself from accountability by manipulating the political process and by undermining the constitutional checks on his power.

In both UDM and EFF 2; the Court invoked this ‘populist’ conception of constitutional accountability to explain why it is appropriate to impose constitutional constraints on the functioning of political parties in the National Assembly where party discipline appears to function as an obstacle to constitutional accountability. In UDM, Chief Justice Mogoeng remarked:

If the will of political parties were always to prevail [mechanisms to ensure executive accountability will often not function]. Conceptually, those majorities could only be possible if members of the ruling party are also at liberty to vote in a way that does not always have to be predetermined by their parties. And this of course assumes that the ruling party will generally be opposed to the removal of their own.¹³⁵

This theme reverberates throughout EFF 2. Jafta Joffered this observation:

¹³⁰ Ibid at paras 6 and 7 (my emphasis).
¹³¹ Arendt (note 32 above) at 50.
¹³² UDM (note 30 above) at para 88.
¹³³ State of Capture (note 21 above) was published on 16 October 2017.
¹³⁴ Ibid at para 142.
¹³⁵ UDM (note 30 above) at para 61.
The fact that members of the Assembly assume office through nomination by political parties ought to have limited influence on how they exercise the institutional power of the Assembly. Where the interests of political parties are inconsistent with the Assembly’s objectives, members must exercise the Assembly’s power for the achievement of the Assembly’s objectives. For example, members may not frustrate the realization of ensuring government by the people if its attainment would harm the political party. If they were to do so, they would be using the institutional power of the Assembly for a purpose other than the one for which the power was conferred. This would be inconsistent with the Constitution.  

This reasoning provides constitutional support for individual members to act independently of their parties by setting constitutional limits to the enforcement of party discipline by the familiar parliamentary practice of ‘whipping’ in service of a fundamental constitutional norm of representative government, as set out in s 1(d) of the Constitution. I have described this norm as a constitutional requirement of principal/agent accountability.

This far-reaching decision constrains not only how parliamentary majorities and the party whips function in the National Assembly, but also the powers of a party’s leadership elected by its members ‘outside’ the Assembly. As a matter of constitutional law, individual members of the National Assembly who act to vindicate a constitutional value against the wishes of their party will now have a constitutional defence to any attempt at removal or ‘punishment’ by their party leaders. This extraordinary outcome was constitutionally justified, I will now suggest, having hedged my bets earlier, by the evidence that ‘party discipline’ was being invoked to protect a corrupt President. He had refused to be held accountable for the misuse of public funds, and a ‘majority party’ absolved him of responsibility for engaging in behaviour that served no legitimate purpose in our constitutional democracy.

How broadly this newly announced constitutional constraint on the functioning of political parties in the National Assembly should be construed remains unclear. Some party discipline is essential to the coherent functioning of political parties as instruments of collective action, electoral representation and the proper functioning of elected National Assemblies. Indeed, the First Certification Judgment Court rejected the argument that the system of party discipline and closed electoral lists was inconsistent with the constitutional requirement of accountability:

We do not agree. Under a list system of proportional representation, it is parties that the electorate votes for, and parties that must be accountable to the electorate. A party which abandons its manifesto in a way not accepted by the electorate would probably lose at the next election.

That said, in both UDM and EFF 2, the Court recognised that in some circumstances, party discipline can also weaken accountability – and the principal/agent relationship between representatives and the electorate. In these two cases it can hardly plausibly be contended

---

136 Ibid at para 144.
137 Certification of the Constitution (note 4 above) at paras 106 and 186.
138 F Rosenbluth & I Shapiro Responsible Parties: Saving Democracy from Itself (2018) 23 (The authors make a persuasive argument that strong disciplined, programmatic parties promote ‘democratic accountability’ (my emphasis). See also J Carey & A Reynolds ‘Parties and Accountable Government in New Democracies’ (2007) 13 Party Politics 255. (The authors distinguish two aspects of the strong party ideal: legislative discipline and programmatic platforms. Consistent with South Africa’s experience, they suggest that legislative discipline can undermine both programmatic platforms and representative accountability to the electorate to see through the programmes for which they were elected). In these cases, ‘party discipline’ was being invoked not to support the governing party’s programmatic objectives but to protect a corrupt President. In these circumstances, the enforcement of judicial limits on ‘party discipline’ was a justified ‘intrusion’ into the ordinary internal functioning of the caucus of a ‘parliamentary party’ and the National Assembly.
that party discipline was invoked to vindicate an electoral mandate. Rather, the cases reflect what Vermeule calls ‘agency slack’. As EFF 1, UDM and EFF 2 make clear, the interests of the electorate in the proper custodianship of public funds was being egregiously ignored by its ‘representatives’, without apparently any prospect of a remedy for a constitutional wrong through the ordinary operation of the representative political process. The majority in the National Assembly acted in their own short-term interests as incumbents and in order to protect a corrupt President. Neither the President nor the majority party could be said in these circumstances to be vindicating electoral promises.

C Principal/agent accountability and the separation of powers

‘Democratic process dysfunction’ in the form of ‘agency slack’ – egregious failures of accountability by representatives to voters – can on occasion then, provide a rationale for judicial intervention. However, it remains the exception, not the rule, and cannot displace standard separation of powers considerations concerning the proper constitutional relationship between the courts and the National Assembly.

In UDM, the Court carefully crafted a decision in which the Speaker had the power to decide whether the vote on a motion of no confidence in the President should or should not be by secret ballot. Her exercise of this discretion, however, had to be appropriately seasoned with considerations of rationality.

A danger lurks. This ubiquitous ‘means/ends’ standard, incorporated into a constitutional law context from the judiciary’s experience in exercising its review powers over the exercise of administrative discretion, does not do the intended work. A secret ballot could just as easily support as undermine the Constitution’s anti-corruption goal. A secret ballot aimed at reducing coercive pressures exercised by party leaders over ‘backbenchers’ could also foster corruption, by insulating individual members from any public accountability.

The Court’s rationality standard will provide little ‘guidance’ on how this conundrum should be resolved by the Speaker in the future without the Court itself hinting at the ‘best’ decision, while allowing for some ‘margin of appreciation’.

Since that might not be the ‘best’ exercise of the Court’s own constitutional authority, two other doctrines might serve as a preferred guide. Principal/agent accountability is one. In just a moment, we explore another.

The separation of power’s question raised by EFF 2 were discussed in part IV. As we saw, the ‘doctrine’ when applied in a very traditional sense raised two concerns. First, the judicial

139 Vermeule (note 16 above).
140 UDM (note 30 above) at para 94.
141 Ibid at para 80. Similar issues arise with respect to voting in elections, which is provided for in s 90 of the Electoral Act 73 of 1998. See S Holmes ‘The Secret Ballot and the Lives of Others’, available at http://www.college-de-france.fr/media/jon-elster/UPL31835_holmes_scrutin.pdf (Concerns the merits of the secret ballot in voting). One side in the argument asserts that the secret ballot protects the voter from ‘bullying pressure’ and the other that the secret ballot permits the voter to ‘indulge his insalubrious motives at public expense.’ The scandal in the UK Parliament some years ago, arising out of the exposé of a parliamentarian who was paid by a wealthy businessman to ask a parliamentary question, should serve as a reminder that secrecy can also corrupt parliamentary proceedings.
intrusion could potentially be ‘disabling’.

Justice Zondo’s dissent noted that that respect should be shown for parliamentary processes by discouraging ‘litigants from approaching this court or any court for that matter in regard to an issue which is capable of being resolved without going to court’. Of course, the Justice meant ‘in the legislature’ itself – or through other political mechanisms, say elections or branch meetings or political assemblies (all of which are exercises in popular sovereignty). Regular judicial intervention opens the door for minority parties to alter the balance of power in legislatures by using the courts as an extra-legislative forum. Such usage could open the courts to charges of ‘partisan adjudication’.

A second institutional consequence of judicial intervention might be described as the ‘Lochner effect’. EFF 2 decided more than the immediate issues in dispute between the litigants on the facts that were before the C. It could have long-term consequences. The impeachment procedure has substantially been constitutionalised by EFF 2. According to the majority judgment, the initiation, determination of ‘jurisdictional facts’, the rules of decision-making, the party-political composition of the Impeachment Committee, and the final outcome are now subject to judicial supervision.

The EFF 2 Court came to these conclusions without considering the sui generis and highly ‘political character’ of the constitutional power to remove an elected President. Indeed, it assumed erroneously that ‘impeachment’ is really no different from ‘confidence’ proceedings. Impeachment proceedings, unlike a confidence question, is the ultimate form of constitutional censure that is only likely to arise as a material question in extreme situations of political crises, fraught with risk to the body politic and the judiciary itself. The Court has now inserted itself in this process, apparently without considering such adverse, long-term consequences.

Precautionary constitutionalism and democratic process reinforcement considerations would have counselled a more limited intervention. It could have simply required the initiation of impeachment proceedings, since the National Assembly was clearly incapable of holding the President to account in the ‘direst of situations’ because of significant democratic process dysfunction. It need not have prescribed ‘constitutional impeachment rules’. When a court intervenes, especially in novel situations, it does not have to decide every question that might arise in the future.

---

142 JB Thayer ‘The Origin and Scope of the American Doctrine of Constitutional Law’ (1893) 7 Harvard Law Review 1, 29. Thayer was concerned that judicial review could sap the democratic process of its vitality and erode its capacity. However, Thayer was writing at a time of greater confidence in the functioning democratic institutions and before judicial review was exercised on a regular basis. Judicial review in contemporary constitutional democracies, should be considered in the light of our experience, including our experience with ‘democratic process dysfunction’. Still, the potentially ‘disabling’ impacts of intrusive judicial review on parliamentary proceedings identified by Thayer, remains important when courts decide on the scope of review.

143 EFF 2 (note 30 above) at para 60.

144 Ibid (note 30 above).

145 Tribe & Matz (note 81 above).

D Towards a principle of accountability

1 Accountability, broadly speaking

Everyone agrees that corruption is evil. But the concept is notoriously difficult to define.\(^{147}\) This article construes corruption broadly and certainly draws ‘political corruption’ within its ambit. When political corruption occurs, the courts have a duty to protect the Constitution as a project of democratic self-government.

The concept of political corruption therefore refers to constitutionally suspect conduct by elected occupants of a public office who have won political power through the electoral process, and who use that power to benefit themselves and their ‘friends’ rather than the people who elected them. So understood, corrupt ‘self-dealing’ representatives violate both their ‘contract with the people’ and their oath of public office: in so doing, they undermine a foundational relationship between representatives and voters upon which our democratic project rests. Such self-serving behaviour by our elected representatives amounts to an ‘abuse of power’, is inconsistent with the ‘rule of law’ and can undermine the constitutional duty to respect, protect and promote the rights enshrined in the Constitution.\(^{148}\) Of course, corrupt politicians usually have partners in crime. As Dennis Thompson notes: ‘If private interests undermine the democratic process to enlist public authority in furthering their purposes, they become agents of corruption’.\(^{149}\)

The Court has increasingly incorporated a principle of accountability into its reasoning. In EFF 1 it acted as the constitutional ally of the Public Protector. The Public Protector had determined that President Zuma had violated s 96 of the Constitution. It cast s 96 in the role of an express ‘anti-corruption provision’. While the EFF 1 Court did make this particular finding, it regularly evoked accountability as a constitutional leitmotif designed to chop off the ‘stiffened neck’ of corrupt officials. In EFF 2, neither accountability nor anti-corruption play an express role in the Court’s express reasoning. Still, the State of Capture report loomed large in the background.

Why is accountability so central to the workings of a principal/agent based account of representative democracy and the more expansive notion of self-governance. As Woolman has written:

[Ongoing] polycentric exchanges between political institutions and social actors that build mutual trust, concern and loyalty in discrete relationships and within informal networks and formal institutions can, cumulatively, create a society and a state identified with a deep commitment to the rule of law [which turns inevitably on the ability to hold others accountable, and ultimately to internalise accountability as a social more]. If these vertical, horizontal and polycentric relationships, networks and institutions are constantly reaffirmed, then we might witness two developments. First, state actors … should enjoy greater legitimacy [as our representative]. Second, the various associations, communities, networks, and sub-publics that constitute civil society – … should be strengthened. On this account, it is clearly false to say that the more formal, purely vertical,

\(^{147}\) Fukuyama (note 3 above) at 83, 84 (Corruption as the ‘appropriation of public resources for private gain’ is a ‘useful starting point’, but it also embraces clientelism, patronage and rent seeking behaviour).

\(^{148}\) Constitution s 7(2).

conception of the rule of law [and the principle of accountability come] first. The relationship is one of reciprocal effect.¹⁵⁰

South Africa, on this account, has witnessed some recent success with the reciprocal effect between the judicial enforcement of (and a broader societal commitment to) the principle of accountability and the foundational commitment to representative democracy:

With that foundation still in place, the fourth estate could break a story in 2009 about malfeasance at the highest level of government: the President’s use of public funds to build a private estate … the Public Protector, could then initiate an investigation. At roughly the same time, the rampant … dysfunctional governance of the South African Broadcasting Corporation became a matter … [for] review … by the Supreme Court of Appeal. … [It] held that the Public Protector’s recommendations and remedies may have legally binding effect. The Supreme Court of Appeal’s holding in *SABC v DA* lit the path for the Constitutional Court’s assessment and reinforcement of the Public Protector’s powers in *Economic Freedom Fighters I* … The two rulings confirmed for many citizens what they had already heard. … The severity of the damage [to the ANC] can be measured by the results of the 2016 municipal elections – just months after *EFF I*. The ANC, for the first time, lost control of virtually all of South Africa’s major metropoles. … The electorate’s judgment had a reciprocal effect on an array of actors. The Constitutional Court felt increasingly confident in wielding its limited powers in matters that turned on accountability … The Public Protector broadened her investigations into the depravity of the President, his patronage system and the non-political actors who benefitted immensely from a blurring of the lines between the public domain and the private realm. Finally, in December 2017, ANC members chose Cyril Ramaphosa as its new party President, and subsequently, President of the nation.¹⁵¹

However, prior to clear evidence of such reciprocal effect, the Court in *EFF I* had already established a judicial doctrine connecting accountability and representative democracy. As the Chief Justice wrote in *EFF I*:

One of the crucial elements of our constitutional vision is to make a decisive break from the unchecked abuse of State power and resources that was virtually institutionalized during the apartheid era. To achieve this goal, we adopted accountability, the rule of law and the supremacy of the Constitution as values of our constitutional democracy. For this reason, public office-bearers ignore their constitutional obligations at their peril. This is so because constitutionalism, accountability and the rule of law constitute the sharp and mighty sword that stands ready to chop the ugly head of impunity off its stiffened neck.¹⁵²

The Court found – in its well established s 1(c)-driven rule of law jurisprudence, along with the principle of accountability in s 1(d) and s 96’s expressly clear warnings regarding ‘self-dealing’ behaviour – sufficient support to hold the President accountable for the systemic damage

---

¹⁵⁰ Woolman (note 37 above) at 374

¹⁵¹ Woolman (note 9 above) at 160–161. See also M Krygier ‘The Quality of Civility: Post-Anti-Communist Thoughts on Civil Society and the Rule of Law’ (at 251) in A Sajó (ed) *Out of and into Authoritarian Law* (2002) 221. Martin Krygier writes that this commitment to the rule of law and accountability provides: ‘[the] … scaffolding for the building of ‘civil’ relations between state and citizens and among citizens themselves. They can begin to rely upon, rather than merely fear, the state and law. Apart from causal relationships, there is … a real affinity between the rule of law [qua principle of accountability] and civil society. Causal links are sometimes hard to trace, but a polity in which both the rule of law [the principle of accountability] has a deep hold is one in which restraint [and respect are] … cultural norm[s] … [They marry well] with pylons firmly planted on both sides of the divide and input moving in both directions.’

¹⁵² *EFF I* (note 30 above) at para 1.
his abuse of the public purse had done to our nascent democracy and the people who rely so heavily on the state to provide desperately needed public goods.  

But the EFF 1 Court’s use of the principle of accountability to buttress representative democracy is hardly new. James Fowkes offers a detailed account of a jurisprudence replete with cases that bind these two linchpins of our basic law together. He begins by noting that “The natural reading of FC s 1(d) envisages the maintenance of “accountability, responsiveness and openness” via the electoral mechanisms of vigorous representative democracy”. His three-part analysis first identifies those cases in which accountability makes representative democracy possible through electoral mechanisms. Consistent with the primary thrust of this article, he notes:

[UDM] (2) therefore articulates a structural understanding in which representative democracy should mostly be allowed to enforce itself, with the judicial role under FC s 1(d) being to enforce the basic ground rules to ensure that the choice of representatives is real and fair.

However, since representative democracy does not, as we have seen, always enforce itself, the Mazibuko v Sisulu Court found that FC s 1(d) plays:

[an] important role in identifying the accountability-enhancing purpose of FC s 102(2), and ultimately in invalidating parliamentary rules that made it virtually impossible for minority parties to have a motion of no confidence debated before the National Assembly.

In a second line of cases, the connection between the broad principle of accountability and principal/agent accountability is somewhat more attenuated. For example, he argues that the Court in Khumalo v Holomisa makes the plausible assertion that ‘the media are important agents in ensuring that government is open, responsive and accountable to the people as

---

153 Constitution s 96 reads, in pertinent part, as follows: (2) Members of the Cabinet and Deputy Ministers may not — (a) undertake any other paid work; (b) act in any way that is inconsistent with their office, or expose themselves to any situation involving the risk of a conflict between their official responsibilities and private interests; or (c) use their position or any information entrusted to them, to enrich themselves or improperly benefit any other person.


155 Ibid at 13–60, discussing United Democratic Movement v President of the Republic of South Africa (No 2) [2002] ZACC 21, 2003 (1) SA 495 (CC) (‘UDM 2’).

156 Ibid at 13–60, discussing Mazibuko (note 61 above). However my different reading of this case concludes that this may be a case of ‘judicial overreach’ on separation of powers grounds (Fn 47 above). What was at issue was whether the majority could delay the scheduling of a motion of confidence. The Court came to the conclusion that rules should be adopted (since there was a ‘lacunae’ in the rules), which treated the scheduling of such motions as a constitutional right. The result is that the timing and scheduling of such motions is now under the control of the government’s parliamentary adversaries. It might be that the Court was of the view that this result would always enhance accountability. But is this so? Suppose the electoral map changes, resulting in unstable coalition governments, vulnerable to displacement between elections through marginal parliamentary votes. Does a rule which enables a small ‘opportunistic’ party to precipitate a change in government at will enhance accountability? In these circumstances, rules might be required that give such coalitions more time to negotiate. Stable government is also an important political value. Yet that possibility is now foreclosed by the Court’s decision to constitutionalise aspects of the ‘confidence procedure’. The Court itself did not rely on the textual principle of accountability in its reading of s 102(2). Rather, it invoked the concept of ‘deliberative democracy’ to justify its decision to intervene in the ‘internal’ functioning of the National Assembly. However, the importance of James Fowkes’ reading lies in his grounding of the interpretation of s 102(2) in the text of the Constitution and the Court’s accountability jurisprudence. After all, we are in search of ‘bright lines’.
the founding values of our Constitution require’.\textsuperscript{157} According to Fowkes, ‘the implication of Khumalo is that FC s 1(d) protects not only the basic mechanisms of representative democracy, in order to achieve accountability, but anything that promotes accountability in a representative democracy, such as a free press.’\textsuperscript{158} However, Fowkes also contends that s 1(d) of the Constitution does not support a free-standing principle of accountability – despite an array of judgments that stand for the opposing proposition.\textsuperscript{159} To find that s 1(d) does so would sever its clear textual connection to representative democracy.\textsuperscript{160}

2 A narrower conception of accountability: a principle of anti-corruption

Should a principle of accountability appear too broad to provide a bright line for judicial intervention in the legislative process, then another narrower option exists: an anti-corruption, precautionary principle. Whilst not as easily rooted in specific textual provisions, there is evidence in the jurisprudence of the Court that can be read in this way.

The earliest is the Heath case.\textsuperscript{161} The Heath Court held that:

[Corruption] in administration is inconsistent with the rule of law and the fundamental values of our Constitution. They undermine the constitutional commitment to human dignity, the achievement of equality and the advancement of human rights and freedoms. And they are the antithesis of the open, accountable, democratic government required by the Constitution. If allowed to go unchecked and unpunished they will pose a serious threat to our democratic state.\textsuperscript{162}

On the Heath Court’s account, political corruption is (a) by definition, a breach of the rule of law; (b) inconsistent with commitment to the Constitution’s human rights provisions; and (c) amounts to a failure of an office-bearer to hold herself accountable to the people she serves, which is a constitutional requirement in a representative democracy. For this C, it functions as an erasure of the foundational values found in ss 1(a) through 1(d) of the Constitution, a panoply of specific textual violations – be they in Chapter 2 (the Bill of Rights) and Chapters 4 to 7 (the structures of representative democracy) – that underpin the judgment.

Similarly, years later, the Glenister\textsuperscript{163} Court found itself confronted with a government that had gone out of its way to destroy an independent institution designed to keep state actors and private actors from engaging in the kind of corruption that violated the rule of law and threatened to degrade democratic institutions. In so doing the President and his cronies freed themselves from oversight and gave themselves a green light to pillage and plunder the public purse. The Glenister Court held that legislation establishing the Directorate for Priority Crime Investigation (DPCI) and disestablishing the Directorate of Special Operations (DSO) because it was constitutionally deficient on the grounds that the DPCI ‘lacks the necessary


158 Fowkes (note 154 above) 13–61

159 Rail Commuters Action Group v Transnet Ltd t/a Metrorail [2004] ZACC 20, 2005 (2) SA 359 (CC), 2005 (4) BCLR 301 (CC) at paras 74–75; S v Mamabolo [2001] ZACC 17, 2001 (3) SA 409 (CC) at paras 37 and 45; K v Minister of Safety and Security [2005] ZACC 8, 2005 (6) SA 419 (CC) at para 23.

160 Fowkes (note 154 above) at 13–65.


162 Ibid at para 4.

structural and operational independence to be an effective corruption fighting mechanism’. The Court’s reasoning relies heavily on South Africa’s ratification of various international law instruments designed to support transnational efforts to combat systemic dishonesty by state actors. But the Glenister Court made it unambiguously clear that this constitutional commitment arose from the text of the Constitution as a whole.

The conclusion that the Constitution requires the state to create an anti-corruption entity with adequate independence is therefore intrinsic to the Constitution itself.

[Leaving] to one side for a moment the Republic’s international law obligations, we consider that the scheme of our Constitution points to the cardinal need for an independent entity to combat corruption. … The Constitution is the primal source of the duty to fight corruption.

Since the Court did not cite a specific textual source for ‘the duty to fight corruption’ it have relied on an implicit constitutional principle derived from a reading of the Constitution as a whole, which includes its ‘structural provisions’ and implicit normative underpinnings. This narrower anti-corruption constitutional principle explains both the Court’s strict review of executive action and parliamentary legislation and its conclusion that both were constitutionally defective. In a number of subsequent cases, anti-corruption considerations played a central role in the Court’s reasoning concerning: the decisions of the Director of Public Prosecutions, the procurement and the payment of social grants, and the funding of political parties. However, in the trilogy of cases discussed here it is the systemic corruption and utter failure of the democratic process to secure accountability that explains EFF 1 and UDM, or operates as the unmistakable background to EFF 2. In each of the former cases, Corruption Watch, an anti-corruption, non-governmental organisation, joined as an amicus. In all three decisions, the Court invoked a requirement of principal/agent probity as the essentialia of democratic self-government – anchored in the Preamble and s 43(2) of the Constitution – in coming to decisions that are not justifiable from a ‘traditional’ separation of powers perspective.

These three cases introduced, for the first time, a ‘democracy protective rationale’ as distinct from a ‘rights protection rationale’ upon which the Court relied in Glenister. These two rationales are not mutually exclusive. Moreover, the anti-corruption principle is not limited to cases of ‘political process dysfunction’. As we know from our rule of law and Bill of Rights jurisprudence, any constitutionally authorised exercise of power – as Michelman and Woolman have shown – would be subject to the scrutiny under the Court’s anti-corruption principle.

---

164 Ibid at para 178.
165 Ibid at para 195.
166 Ibid at para 200.
167 Ibid at para 175 (my emphasis).
169 Black Sash Trust v Minister of Social Development [2017] ZACC 8, 2017 (3) SA 335 (CC).
172 Glenister (note 163 above) at para 177.
VI CONCLUSION: DEMOCRACY AND INTEGRITY

The Court has now intervened in a significant number of cases that set rules for the political process. Once one overcomes anodyne expressions of the separation of powers doctrine, the departure from ‘well established’ understandings of the proper relationship between the judiciary and other branches in constitutional democracies begins to make sense in South Africa. That said, our fundamental commitment to representative democracy should urge some caution when it comes to the judicial intervention in the ‘internal functioning’ of representative branches, especially with regard to rule-making.

This article has put forward a theory of ‘democratic process-reinforcing review’. Given the deep commitment of the basic law to self-government in multiple forms, this theory holds that where self-dealing representatives: (a) fail to fulfil their custodial responsibilities for public resources on behalf of ‘the people’; (b) engage in self-dealing; (c) are ‘captured’ by private interests; and (d) create corrupt patronage systems that threaten the entire democratic project, then the Court is well within its rights to impose binding legal constraints on the political process in order to secure accountability.

Of course, where accountability ends, and meddling starts, can sometimes be difficult to ascertain. On several occasions the Court has failed to offer a persuasive argument for intervention in the internal processes of representative branches of government. But while the outcomes might appear jarring on a first read, this article has shown that the Court’s implicit recognition of a constitutional requirement of ‘principal/agent probity’ in a representative democracy, which bears a reciprocal relationship with the express principle of accountability in a constitutional state, provides both justification for the more troubling decisions reached, and bright lines for future cases. The theory of ‘democratic process-reinforcing’ review propounded here enables us to ascertain, as best as we can, when the rough and tumble of democratic politics – as a deep constitutional commitment – must be respected and when the democratic process has become so utterly dysfunctional and our representatives so remiss in their duties as ‘representatives of the people’ that our representative democracy is rendered ‘representative’ and ‘democratic’ in name only. Then, when the Court imposes legal constraints on the democratic process in order to ensure accountability, and fidelity to the electorate, it acts on behalf of the people, even though it is not itself elected.
Plea Bargaining in South Africa: An Economic Perspective

RICHARD ADELSTEIN

ABSTRACT: This essay applies a simple economic model of the plea bargaining process to the two-tiered structure of negotiated pleas and sentences in South Africa. Bargaining in South Africa proceeds along one of two tracks. A formal procedure, authorised and regulated by s 105A of the Criminal Procedure Act, gives defendants represented by counsel access to precise information about the terms of the bargain before the plea is made and permits them to withdraw their pleas should the sentencing judge reject the agreement. In contrast, an older, informal procedure, governed by s 112, applies to defendants without counsel and grants them significantly less information and agency in the bargaining process than does the s 105A procedure. The model illuminates the central role of information and uncertainty in the defendant’s decision to plead guilty or insist on a full trial, and suggests that if all plea bargains were governed by the formal procedures of s 105A, the system would more effectively represent the interests of both prosecutors and defendants by producing more bargains, and fewer trials, in cases where both sides want to avoid trials and consummate plea bargains. The final section considers the constitutionality of plea bargaining under ss 35 and 36 of the Constitution of the Republic of South Africa. It reviews the constitutional history of plea bargaining in the United States to emphasise the differing perspectives on the constitutionality of plea bargaining demanded by significant variations in substance and interpretative style in the two constitutions. The essay concludes by briefly suggesting the arguments that might be made against the constitutionality of plea bargaining under s 35 and the corresponding contentions that might be raised during limitations analysis under s 35 to justify the practice should it be found to violate s 35.

KEYWORDS: guilty pleas, sentence agreements, formal/informal plea bargains, information, uncertainty, s 35 right to trial, limitations of s 35 right to trial

AUTHOR: Woodhouse/Sysco Professor of Economics, Wesleyan University. Email: radelstein@wesleyan.edu

ACKNOWLEDGEMENTS: I am grateful to Stu Woolman, Bryant Jackson-Green, and two anonymous referees for helpful comments on an earlier draft. Remaining errors are my own.
I TWO TRACKS TO AGREEMENT

A negotiated guilty plea, or plea bargain, takes place when a defendant agrees to forego his right to a full criminal trial by pleading guilty to some offence in exchange for a favourable concession by the prosecution or, less frequently, the court. Where prosecutors exercise substantial discretion in the selection of charges, the range of possible concessions is wide, particularly where the crime is complex or involves several perpetrators. To secure a guilty plea in the circumstances of the case, prosecutors might, among other things, agree to reduce or to drop the charges faced by the defendant; a co-defendant or some third party, protect the defendant from imprisonment or other onerous sanctions; limit the evidence or facts that might be led or revealed to the court; or stay their hand in some other way that might induce, or pressure, defendants to surrender their right to trial.1 However, in practice, in the many and varied forms of plea bargaining that have emerged in recent decades in both adversarial and inquisitorial systems around the world,2 the concession ultimately sought by almost every defendant is a reduction in the fine or prison sentence that he will receive upon conviction, punishments whose relative severity can meaningfully be represented as objective numbers. In these cases, a defendant pleads guilty to some offence in exchange for a measurably more lenient sentence than he expects to receive after conviction at trial. For the better part of a century plea bargains of this kind have accounted for at least 90 per cent of all criminal convictions in the state and federal courts of the United States.3 To focus attention on the essential aspects of the exchange, I assume here that every plea bargain takes this straightforward form.

Guilty pleas have been negotiated in South Africa for several decades.4 After the passage of the Criminal Procedure Act 51 of 1977 (‘CPA’), what came to be called ‘informal’ plea bargains appear to have been routinely concluded under s 112 of the CPA, which governs guilty pleas at summary trials. Section 112(1)(a) provides that, where a defendant pleads guilty to an offence that does not merit imprisonment or a large fine, and the prosecutor has accepted the plea, the presiding judge may convict the defendant of the offence to which he has pleaded and impose any lawful sentence, without the taking of any evidence beyond the guilty plea itself. For more serious offences involving imprisonment or large fines, s 112(1)(b) still allows the judge to convict and sentence a defendant without trial, provided that she questions the defendant in court to establish a factual basis for the plea, an admission by the accused of facts sufficient to prove the charge to which he has pleaded guilty. While s 112(2) provides for the submission of a written statement of admission in lieu of the oral colloquy, the judge remains free to question the defendant in court in addition to receiving the statement. Nothing in the

section explicitly contemplates a plea bargain, and while the prosecutor may recommend a specific sentence to the court, the judge retains full discretion either to reject the plea under s 113 and force the parties to a full trial, or accept the guilty plea, convict the defendant, and impose any lawful sentence for the offence to which the plea has been entered. The defendant does not hear the sentence until the plea has been entered and the factual basis established. At this point, the defendant may no longer withdraw the plea and demand a full trial.5

Despite its apparent indifference to the practice, once the final procedural piece is in place, s 112 becomes, as Uijjs AJ put it in 1999, ‘virtually tailor-made for plea bargains’.6 This was provided in 1985 by State v Ngubane.7 The defendant was charged with murder and, after discussions with the prosecutor, pleaded guilty under s 112 to the lesser offence of culpable homicide. The presiding judge nonetheless convicted Ngubane of murder with extenuating circumstances. On appeal the court overturned the conviction, reading ss 112 and 113 together to conclude that:

[O]nce the defendant has entered a guilty plea that the prosecutor agrees to accept, the prosecutor’s acceptance limits the ambit of the lis between the state and the accused in accordance with the accused’s plea. . . . That the lis is restricted appears from ss 112 and 113. The proceedings under the former are restricted to the offence “to which he has pleaded guilty” and the latter must be read within that frame.8

This established holding ensures that, irrespective of what the defendant has actually done and what charge might accurately describe it, once the prosecutor and the defendant have agreed on the charge to which the latter will plead guilty, and the defendant admits facts sufficient to support a conviction on that charge, the presiding judge is bound either to reject the plea or sentence the defendant in accord with the statutorily authorised punishments for the offence to which he has pleaded guilty. Even if Ngubane had actually committed murder, that is, once the prosecutor agreed to accept his plea to culpable homicide and Ngubane admitted ‘facts’ sufficient to support this lesser charge, however true or false those admissions might be, he could no longer be sentenced to more than the maximum sentence for culpable homicide on his plea. In this way, the prosecutor can control the maximum sentence to which the defendant will be liable by agreeing to accept a guilty plea for an offence less serious than what he may believe the defendant actually committed. If the difference between the penalty for murder and that for culpable homicide is, say, fifteen years in prison, a prosecutor can offer a defendant whom he thinks is actually guilty of murder the chance to plead guilty to culpable homicide. ‘Save me the expense and uncertainty of a trial for murder’, he says to the defendant, ‘and I’ll save you fifteen years’.9

Given the heavy caseloads faced by criminal courts in South Africa,10 it is not surprising that bargains of this sort quickly became common. In North Western Dense Concrete CC v Director

---

6 North Western Dense Concrete CC v Director of Public Prosecutions (Western Cape) 1999 (2) SACR 669, 677c (C) (‘North Western Dense Concrete CC’).
7 State v Ngubane 1985 (3) SA 677 (A) 617, 683. Cf Clarke (note 4 above) at 164–165.
8 Cf Clarke (note 4 above) at 160–162.
9 Cf Clarke (note 4 above) at 160–162.
10 Steyn (note 1 above) at 207; de Villiers (note 5 above) at 245; Clarke (note 4 above) at 142.
of Public Prosecutions (Western Cape), a case to which we will return, the High Court bowed to the inevitable. The court not only acknowledged the pervasiveness of plea bargaining in South African courts but, where the Director of Public Prosecutions (DPP) had reinstituted charges against the applicant despite a plea agreement that had provided that they would not be brought, enforced the terms of the bargain and ordered the charges dismissed. Though he was not ‘filled with joy’ to say so, Uis J recognised that plea bargains had become an ‘entrenched, accepted and acceptable part’ of South African law, taking place daily at every level of criminal justice. He compared them to civil settlements, which also adjust the *lis* to the needs of the parties and whose validity is unquestioned, and suggested that the entire system of criminal justice might break down if the courts refused to enforce individual bargains or struck down the practice altogether.11

Despite the High Court’s resigned legitimisation of the practice, uncertainty as to the legality of plea bargains remained. On its face, s 112 is about guilty pleas, not negotiations and agreements over sentencing. It enables the bargaining system itself to remain underground, hidden from public view; indeed, had one of the parties not broken the agreement in *North Western Dense Concrete*, the case and the bargain at its core would never have been subject to judicial scrutiny at all.12 Soon after this decision, as part of a larger inquiry into South African criminal procedure, the South African Law Commission (as it was then called) issued a report recommending that the CPA be amended to provide explicitly for plea bargains, and thus exchanges of guilty pleas for leniency in sentencing. So they could be regularised and regulated, the South African Law Commission offered a statute embodying its proposals for legislators to consider. With some modifications, this recommendation was promptly followed. The amended CPA contained a new s 105A – governing plea and sentence agreements.13

Section 105A is a lengthy provision. It regulates its subject in minute detail, ostensibly to protect defendants from overzealous prosecutors who might deprive them of their right to a full trial. In many respects, it is a model statute. It allows only prosecutors who are authorised by the National Director of Public Prosecutions to do so to negotiate agreements with defendants, who must be legally represented. Before entering into an agreement, the prosecutor must consult both the investigating officer in the case and the complainant, advising the latter on the possibility of compensation as part of the agreement. Any agreement on plea and sentence must be reached before the defendant is asked to plead, and the specific terms of the agreement, the charge to which the defendant will plead, the facts to be admitted by the defendant in support of the plea, and the sentence to be imposed by the court after the plea, must all be made explicit in a detailed written document, signed by both parties, that affirms that the defendant’s agreement was given voluntarily and in full knowledge of the rights that he would surrender with the guilty plea. The agreement is then put before the presiding judge, who may hear evidence and take statements from the defendant or others before passing sentence. Once all these procedural requirements have been met, the judge, nominally still in full possession of the authority to sentence, considers the sentence proposed in the agreement. If she finds the sentence to be ‘just’, she then informs both sides of this finding, convicts the defendant on the guilty plea, and imposes the sentence provided in the agreement. If she finds the sentence

---

11 *North Western Dense Concrete CC* (note 6 above) at 683f, 674e, 676f and 678c.
12 Rodgers (note 5 above) at 242–243.
13 South African Law Commission ‘Simplification of Criminal Procedure (Sentence Agreements)’ (2001). See also Rodgers (note 5 above) at 241; de Villiers (note 5 above) at 244; Criminal Procedure Second Amendment Act 62 of 2001, s 105A.
AN ECONOMIC PERSPECTIVE ON PLEA BARGAINING IN SOUTH AFRICA

‘unjust’, either too harsh or too lenient, then she must inform the two sides of what sentence she would consider just, and offer them the chance to revise the bargain on her new, preferred terms. If either side objects to the judge’s terms, then they may withdraw from the agreement, after which the defendant is entitled to a new trial before a different presiding judge.\(^{14}\)

In a perfect world, as discussed below, CPA s 105A might operate to guarantee all defendants a procedure for the inducement of guilty pleas that ensures that every plea is made fully voluntarily, in the defendant’s own interests, and with full knowledge of the sentence to be imposed upon the plea and the specific consequences that flow from the surrender of the right to trial. But it is easy to see how the imperfect world of criminal justice might frustrate this purpose. For one, s 105A applies only to defendants able to afford legal representation, or to whom representation is given. Sections 35(2)(c) and 35(3)(f)–(h) of the Constitution of the Republic of South Africa, 1996, guarantee legal representation at state expense to defendants who are detained and those being tried ‘if substantial injustice would otherwise result’. But fulfilling this promise even for these defendants has proven difficult. And it is not clear that it extends to defendants not under detention in the period between arrest and trial, when most plea bargaining takes place. Since 1996, South Africa has struggled to provide representation to poor defendants in the face of very high rates of crime, a lack of trained attorneys, inadequate funding and poor administration of the state’s legal aid services.\(^{15}\) The great majority of defendants are thus unable to avail themselves of the protections of the formal procedure, and must rely, as they always have, on the informal procedure of CPA s 112.

On the other hand, the time- and labour-consuming requirements of formal authorisation from supervisors and extensive consultation with victims and the police often deter harried prosecutors facing unmanageable caseloads from employing the s 105A procedure when s 112 proceedings will do equally well. While there is evidence that s 105A procedures are being employed in some courts, the effect of these conditions, as Steyn observes, has been to confine bargaining in the great majority of cases to the informal, shadowy realm of s 112 rather than allowing plea agreements to migrate to the formal, rule-governed domain of s 105A. This confinement frustrates precisely those reforms the new law was designed to implement.\(^{16}\) The same pressures of time and material cost that propelled the replacement of costly trials by cheaper plea bargains in the first place continue to bear heavily on South African criminal justice, forcing the replacement of more costly forms of plea bargaining with cheaper ones.

The result is a two-tiered system of plea bargaining in South Africa, two alternative tracks to negotiated sentences on which passengers travel first- or second-class. The first-class track is governed by s 105A, with its close supervision of prosecutors by administrative superiors, its attention to the position of the police and the interests of victims, and the care it takes to ensure both that defendants are fully informed of the precise terms of the agreement and the rights they will surrender by pleading guilty and that their consent to the agreement is given freely. But travel on this track is very expensive, not just for defendants, few of whom can afford the legal representation that s 105A makes the ticket for first-class travel, but for the state as well.

\(^{14}\) Criminal Procedure Second Amendment Act 62 of 2001, s 105A. See also de Villiers (note 5 above) at 245–250; Rodgers (note 5 above) at 244–255.


which must make good the prosecutorial time and effort required to certify a formal bargain. As a result, the great majority of South African defendants must travel second-class, on the informal track governed by s 112. Here prosecutors operate with far less oversight than on the first-class track, the information defendants have at the moment they must decide how to plead is significantly less complete and reliable, judges retain a freer hand in sentencing in cases where an agreement has been reached, and pleas remain binding even where the information on which a defendant has relied turns out to be faulty.

My purpose here is to examine the operation of this system by considering it in light of some very simple economic logic. If there is any legal institution that might usefully be viewed from an economic perspective, one might think, it would be plea bargaining, the forthright exchange of valuable concessions between parties who each stand to gain from the trade. I hope to suggest more precisely how this might be true, but not by assaulting the problem with complex theory or mounds of statistical data. Instead, I propose a simple, intuitive, empirically plausible depiction, a model, of how plea bargaining works that allows us to see more clearly how the behaviour of prosecutors, defendants and judges and the existence of various conditions in the bargaining environment affect the system’s outcomes, the individual agreements that result in guilty pleas. It identifies the factors that move both prosecutors and defendants to seek bargains, shows how these factors affect their decisions, and examines the problems bargainers face when the information upon which they must base their decisions is unreliable or unavailable. It illuminates the differences between formal plea bargains under s 105A and informal bargains under s 112, showing how imperfections in the latter environment can lead to dysfunctional outcomes that are less likely to occur in the case of s 105A bargains. To the extent it succeeds, the model may be a valuable conceptual tool for South African scholars and policymakers, a fruitful way to understand plea bargaining, the problems it solves for the criminal process, and the new problems it creates at the same time.

But it is important to be clear as to what analysis of this sort can and cannot do. Though it does enable us to ask more salient questions about how the plea bargaining system operates, how to improve its outcomes, whether on balance it is a good or a bad thing, and what its constitutional implications might be, it cannot answer many of these questions at all, and can answer others only in a conditional or qualified way that leaves ample room for disagreement on political, moral or jurisprudential grounds. Economics, like every social science, has both a positive and a normative dimension, questions about what is and questions about what ought to be. The former suggests that economics is a science, a way to understand how the social world actually works, how relationships between individuals, groups and institutions are in fact conducted, and how social outcomes that we can see are produced or created. In practising positive economics, economists are like physicists. They try to provide answers to questions about how things really are in terms that all rational people are bound to accept, just as they are bound to accept that the earth is round, whether they like it or not. But physicists are typically not concerned with the rights of electrons or asteroids, or whether it is a good thing that force and mass are related as they are, or whether we should try to change the laws of thermodynamics because they lead to unjust results. All physics is positive science; it does not ask, or answer, normative questions.

The subject matter of economics, in contrast, is living people and their relations to one another. In this realm of perpetual political and moral controversy, normative concerns, ancient, ultimately unresolvable questions about what justice demands, whether the way things
are is the way they ought to be, and what the object of public policy, or social life itself, should be, irrepressibly jostle with positive questions for our attention. Positive economics can inform normative debate, by showing how and with what consequences various normative positions and the policies that flow from them might play out in practice. But positive economics alone can never answer or determine normative questions. How plea bargaining works, how its outcomes are produced, and how various conditions affect its operation are positive questions, ones about which economic reasoning may have much of value to say. What it would mean to make the plea bargaining system work ‘better’, or whether it should exist at all, are normative questions, and in addressing them, economists have no claim to greater wisdom or expertise than anyone else. But if it cannot answer them, economic reasoning may serve to sharpen these questions, to show just what is at stake and how normative or constitutional concerns can be identified and addressed, and with what consequences. I hope here to show the value of an economic perspective on plea bargaining in both these contexts.

In part II, I start from the assumption that plea bargaining is a desirable, or at least uncontroversial, institutional innovation that raises no normative or constitutional issues. This approach makes it meaningful, after sketching how the system actually works, to ask how to make it work better, a question sharply posed by the two-track South African system. With the help of a general overview of the plea bargaining system and a simple algebraic depiction of the defendant’s problem in deciding whether to plead guilty or insist on a trial, I argue that bargaining under s 112 is fraught with problems stemming from uncertainty. Because defendants do not, or do not always, have accurate or reliable information about the factors that most strongly influence their decision to plead guilty or insist on a trial, the plea bargaining system is likely to fail in a specific way. Plea bargains that would have been made had defendants had the information that they needed will not be made, and plea bargains that would not have been made had defendants had this information will be made, because they do not. ‘Making the system work better’ thus means facilitating the former type of bargain and impeding the latter, a problem of institutional design and performance. If a criminal process commits itself, reluctantly or enthusiastically, to plea bargaining as a legitimate means of resolving criminal cases, it ought to prefer a bargaining system that works well at identifying and completing genuinely consensual agreements to a system that works poorly. I argue that the procedure of s 105A, cumbersome as it is, addresses most of the informational problems that arise under s 112. That is, cost considerations aside, the system’s operation would be improved – more bargains that should take place, and fewer that should not, will take place – if all South African plea bargains were formal and governed by s 105A.

The superiority of s 105A bargaining to its informal counterpart on various grounds has been widely observed. While the courts have as yet had nothing to say on the question, several scholars have noted that this disparity raises a serious problem under s 9 of the Constitution, which guarantees equality before the law. The majority of defendants, who are poor, cannot take advantage of the formal protections and guarantees of s 105A because they, unlike wealthier defendants, generally have no legal representation.17 In the final part of this article, following

---

17 De Villiers (note 5 above) at 253–255; Steyn (note 1 above) at 217–218; Rodgers (note 5 above) at 261; A Borman An Evaluation of the Benefit of Plea and Sentence Agreements to the Unrepresented Accused (2016, unpublished LLM thesis, Faculty of Law, University of the Western Cape) 82–91. A rather different array of jurisprudential objections to plea bargaining are discussed in M Bennun ‘Negotiated Pleas: Policy and Purposes’ (2007) 20 South African Journal of Criminal Justice 17. I consider some of these objections below.
these scholars, I relax the working assumption of part II and ask not whether plea bargaining is a good thing, but how it comports with the structure and interpretive values of the South African Constitution. A brief look at the constitutional history of plea bargaining in the United States raises a range of normative questions about the system of plea bargaining rather different from the one that concerned these scholars, and engages the interpretive strategies employed by the American federal courts to address them. I then compare these constitutional questions and interpretive strategies to the alternative approaches suggested by the particulars of South African law and institutional development, and conclude with a brief sketch of the issues that might frame a constitutional challenge to plea bargaining in South Africa. I hope in both these ways to advance the important emerging debate over plea bargaining in South Africa.

II MAKING PLEA BARGAINING WORK BETTER

A How plea bargaining works

Challenging Judge Uijss’s analogy of plea bargaining to civil settlements in North Western Dense Concrete, Mervyn Bennun called the differences between the two so sharp that ‘the comparison verges on the cynical’. But it was, he went on, ‘useful’ in pointing out the resemblance of plea bargaining to ‘the purchase of a rug at a Lebanese bazaar’:

To move away from the outcome of a criminal trial as being unpleasantly redolent of the marketplace, a plea and sentence agreement negotiated under s 105A should result in the same verdict and sentence which would have followed had the matter gone to a full trial, and the court should satisfy itself that this is the case. . . . [A] negotiated agreement under s 105A [surely] should propose the sentence which would have been imposed after a full trial. 18

Perhaps Bennun, a sharp critic of the practice, proposed this interpretation knowing that it would, if adopted, largely eliminate plea bargains under s 105A and render them available only under s 112. Under s 112, he thought, leniency might still be justified if the presiding judge were satisfied that the defendant’s plea showed genuine remorse. 19 But his interpretation of s 105A plea bargains would render them pointless, and thus make them disappear. 20

The crux of every consensual plea bargain is the sentencing discount given in exchange for the guilty plea. As elaborated below, if there is no difference between the sentence the defendant would receive after conviction at trial and the sentence proposed in the agreement, then there is little reason for a defendant with even a very slim chance of acquittal to accept an agreement, convict himself without putting the prosecution to the proof, and surrender his right to take the small chance that he will somehow avoid conviction. Even if the plea is motivated, as some are, not by fear of greater punishment after trial but by the adverse consequences of bringing the facts of the matter to light in a public trial, s 105A’s requirement that the facts be stipulated in writing largely removes this advantage of guilty pleas.

As Bennun suggests, the exchange of sentencing discounts for guilty pleas is not pretty, but if the system of plea bargaining is to work at all, this is the way it has to work. Everywhere it has taken hold, plea bargaining has been an institutional response to the needs of the state, not the interests of defendants. When prosecutors and administrative judges face caseloads so heavy that only a small sliver of cases can possibly be afforded a full trial given the resources at hand,

18 Bennun (note 17 above) at 29.
19 Ibid.
20 Compare Watney (note 16 above) at 226.
an environment in which the American criminal process has laboured since at least 1900\textsuperscript{21} and, as Steyn suggests,\textsuperscript{22} has characterised South Africa for at least several decades as well, prosecutors have little choice but to somehow persuade large numbers of defendants to surrender their right to a trial and agree to convict themselves. This deeply entrenched process obviates the need to mount a trial in those cases and allows the state, as the provider of criminal justice, to put its extremely scarce prosecutorial and judicial resources to better use elsewhere. Absent the public expenditure necessary to provide a full trial for every defendant, the alternative, as Uijjs J recognised, is the collapse of the system itself. Because so few defendants are so remorseful or so unwilling to endure the publicity of a trial to agree to plead guilty without some material inducement – something they want and deem sufficient to compensate for the loss of rights and the possibility of acquittal their plea entails – prosecutors must offer something in return for the guilty pleas they need to keep the criminal process afloat. In plea bargaining systems everywhere, sentencing discounts are the only answer.

One way bargains might be achieved, sentence bargaining, would have the defendant or his legal representative bargain directly with the sentencer, the trial judge. But concerns about judicial neutrality and dignity, in a process most observers (and participants) see as demeaning, keep judges in almost every adversarial system, including South Africa and the United States, from intervening directly in plea negotiations. The alternative, charge bargaining, is almost universally employed because it is so well adapted to a peculiar feature of the adversarial system, one that prosecutors facing severely overcrowded dockets exploit to the fullest. In inquisitorial systems, like those of continental Europe, prosecutors are generally bound by a rule of compulsory prosecution. This type of rule requires them to try every case on the highest possible charge the evidence will support, nothing more or less, so as to produce a judgment that, to the greatest extent possible, accurately reflects the actual facts and moral weight of the case. If a defendant, like Ngubane, is actually guilty of murder, then the prosecutor must charge him specifically with and try him for murder.

But subject to a provision for private prosecution, at common law and under s 6 of the CPA, South African prosecutors have almost complete discretion to decide whether or not to prosecute any defendant and, if so, for what crimes.\textsuperscript{23} At the same time, the substantive criminal law establishes a large repertory of crimes covering a broad range of behaviours, most modified by degrees, many overlapping or governing activities typically undertaken as part of a single criminal transaction.\textsuperscript{24} This profusion of distinct offences gives prosecutors a great deal of sentencing power. Subject to whatever constitutional or administrative prohibitions of vindictive overcharging might exist, a prosecutor can typically charge a suspect with several independent crimes committed in the same course of action, exposing him to a potential punishment much greater than that prescribed for the principal crime in the case. Most importantly, as in Ngubane, since there is no rule of compulsory prosecution, the charge the prosecutor selects need not accurately describe what she believes the suspect has actually done. So she can charge a suspect with any crime or combination of crimes on the list that he has committed, or with some lesser offence that he has not actually committed. This latitude


\textsuperscript{22} Steyn (note 1 above) at 207–208, 218.

\textsuperscript{23} Criminal Procedure Act 51 of 1977, s 6; Steyn (note 1 above) at 206–207.

\textsuperscript{24} G Kemp (ed) Criminal Law in South Africa (2nd Ed, 2015).
forces the judge to sentence the defendant to the punishment the prosecutor has promised in the bargain.

Plea bargaining in this form is made possible by two aspects of adversarial procedure that are absent in inquisitorial systems. First, defendants have the right to plead guilty and abort trials before they occur. Section 35 of the Constitution guarantees every accused the right to a fair trial, but it does not command anyone to exercise it. Defendants may have reasons other than, or in addition to, an offer of leniency to avoid a trial, and pleading guilty allows them to do it. A guilty plea is a conviction, an agreement by the defendant to submit to punishment for the offence to which he has pleaded guilty, and it ends the criminal case. Typically, the only consequence of the plea to the defendant is the sentence to which he has agreed. Inquisitorial defendants cannot plead guilty or abort their trials, because the point of inquisitorial trials is to provide an accurate account of what actually happened and establish an appropriate punishment for what the perpetrator has done. But adversarial trials, adorned as they are with complex rules of evidence and procedural rights, like the right against self-incrimination, designed to protect defendants from the consequences of the truth, are not primarily about discovering the truth of the matter. They are, ideally, about finding guilty defendants criminally liable for something that allows appropriate punishments to be imposed on them. So the right to plead guilty, the power to subject the prosecutor to a costly trial or make it unnecessary, becomes a valuable bargaining chip for defendants in negotiations. Defendants have something to give that prosecutors want.

The second enabling feature of the adversarial system is that prosecutors, because of their broad charging discretion, have something of value to give them in return. By selecting charges and combining them with sentencing recommendations that, at least in the United States, they know judges, who operate under the same pressures they do, will respect, prosecutors can calibrate punishments as finely as necessary to win guilty pleas from defendants with penalties they consider in each case to be acceptable discounts for the defendant’s saving the government the cost and uncertainty of trying them. ‘Save us this much money’, the prosecutor says, ‘and we’ll save you this much time’. Again, this requires that defendants, in the last analysis, be the ones who bear the state’s costs of convicting them at trial, in the form of heavier punishments upon conviction at trial, so as to preserve their incentive to bargain. Plea bargaining is only possible where defendants who do insist on a full trial are consistently sentenced upon conviction with more severe punishments than they would have been had they accepted a plea agreement. In this way, prosecutors can make surrendering the right to the trial attractive to all defendants. Only then will a proffered sentencing discount be seen by defendants as credible.25

Where defendants are denied their freedom before trial, and the direct and collateral consequences of detention weigh heavily on them during bargaining, the prosecutor’s hand is greatly strengthened, and the threat of prolonged pre-trial detention can be employed strategically to induce pleas from defendants on terms more favourable to the prosecution than would otherwise be the case. Much the same effect is created by statutes that attach mandatory minimum sentences to conviction for specific offences. The political pressures that produce these statutes generally ensure that the minimum sentences are relatively harsh compared to the perceived inclinations of sentencing judges. If so, and judges apply these elevated sentences consistently, without ‘nullifying’ them by refusing to find defendants guilty of these offences

because judges believe the statutory minimum is inappropriate, then mandatory sentencing statutes allow prosecutors to threaten defendants with very severe punishments if they are convicted of these crimes at trial. As we will see, this increase in the punishments imposed after conviction at trial means both fewer trials at which to impose them, because more bargains will be struck by defendants eager to avoid trials that might result in the statutory minimum, and bargains more favourable to the prosecution. In other words, mandatory minimum sentences will extract guilty pleas from defendants to lesser offences without mandatory minimum sentences at punishments greater than would have resulted without the minimum sentences in the background. In this way, the effects of the mandatory sentences ramify throughout the criminal process, increasing punishments for the crimes to which they do not apply as well as those to which they do.26

This much is true of every mandatory sentencing law. But South Africa’s mandatory sentencing provisions, s 51 of the Criminal Law Amendment Act 105 of 1997 (‘CLAA’), intended to be a temporary response to a perceived wave of violent crime across the country,27 add a further twist. The statute makes life imprisonment mandatory for defendants convicted of murder and rape under specific aggravating circumstances, mandates fifteen-year sentences for a range of other crimes, requires sentences from fifteen to twenty-five years for repeated offences in various situations, and allows judges to impose lesser punishments only where they are ‘satisfied that substantial compelling circumstances exist which justify’ them, though that exception was construed quite narrowly in State v Malgas.28 However, the mandatory sentencing provisions apply only to cases tried in a High Court or a regional court. They do not bind district magistrates’ courts, where most cases involving offences listed in ss 51(1) and 51(2) are tried. So in addition to the bargaining power conferred by their broad discretion in selecting charges, sometimes further enhanced by extensive pre-trial detention of the defendant, prosecutors can increase the pressure on defendants charged with crimes subject to s 51 by threatening to try their cases in a High or regional court rather than a district court. All of this, combined with the general intrusion on the traditional prerogatives of sentencing judges represented by the statutory mandates, has made s 51 deeply unpopular among South African judges and commentators.29 But their disapproval has not dislodged it; the temporary legislation was upheld by the Constitutional Court in State v Dodo (2001) and extended every two years until 2007, when it was made permanent by the Criminal Law (Sentencing) Amendment Act 38 of 2007.30

The institutions of adversarial procedure make plea bargaining possible. What makes it necessary in the systems that have come to rely on it is a combination of two factors common to most of them: (i) the sheer volume of crime and correspondingly large caseloads these

28 Criminal Law Amendment Act 105 of 1997, s 51(1–3); Terblanche (note 26 above) at 197–215; State v Malgas 2001 (2) SA 1222 (SCA).
29 Van Zyl Smit (note 27 above) at 205-208; Terblanche (note 26 above) at 218-220; Cameron (note 27 above) passim.
chronically underfunded systems must confront, and (ii) the elaboration of the trial process itself, the complexity of many modern statutes, the difficulties of proof they pose, the evolution of procedural safeguards for defendants and rules of evidence that increase the length, cost and uncertainty of full trials. The proportion of convictions in a system that result from plea bargains, 90 per cent in the United States, and the extent of the sentencing discounts that must be offered to secure the needed rate of guilty pleas, are measures of the pressure exerted by these factors on the system. In South Africa, though trial procedures are simpler than in the United States, the incidence of crime is comparable. As a result, prosecutors and courts still do not have sufficient resources to meet the demands placed on them. The absence of data makes the effect of the above factors harder to determine. However, if Uijs J is right, and the South African criminal justice system, like its American counterpart, could not withstand the abolition of plea bargaining, the practice must be very widespread, and the discounts correspondingly deep.

B The defendant’s problem

Although the prosecutor may control the charge and, to a large extent, the sentence, there can be no bargain without the defendant’s consent. Thus, if guilty pleas are entered knowingly and voluntarily, the defendant controls the plea bargain. Prosecutors for whom inducing a guilty plea in a given case is not especially urgent, perhaps because they have some special reason to take the case to trial, or because the volume of crime generally does not press too heavily on their resources, may offer only low prices for a plea. Put another way, defendants in these cases will receive smaller sentence discounts relative to the sentences these defendants could expect to receive after conviction at trial. Where, as in the United States, the need to induce pleas is acute because there are far too many cases to try with the available resources, prosecutors may offer very high prices for guilty pleas. Such deep discounts make the negotiated sentence a mere fraction of the expected sentence after a trial. But whether the proffered discount is high or low, it is the defendant who must agree to accept it. Only he can make the trial unnecessary.

So it is useful to examine the defendant’s problem in a plea bargain more closely. Should he accept the sentence proposed by the prosecutor, or should he retain his right to trial by refusing the bargain and pleading not guilty? On the one hand, we will assume for now that the bargain represents a sentence the defendant can count on, that he knows for certain. On the other, going to trial presents the prospect of two possible outcomes, an acquittal, which for simplicity we will represent as a punishment of zero, and a conviction, which brings a punishment that, again for now, we will assume the defendant knows, along with some perceived likelihood of one or the other. Suppose, for example, I am accused of aggravated robbery, for which, if I am convicted at trial, I am certain to receive a ten-year sentence. My defence is that I did not do it, which may or may not be true. But there is evidence against me: my gun was used in the robbery, and my fingerprints were found on it afterward. All things considered, my attorney tells me, the chance of my being convicted at trial is 70 per cent. So the expected punishment I face in going to trial, defined as the sentence that I would receive on conviction multiplied by the likelihood that I will in fact be convicted, is $0.7 \times 10 = 7$ years. This calculation does not mean that I will be sentenced to seven years, since the only possible outcomes of my trial are ten years or none. It means that if, hypothetically, my trial could be conducted identically a million times before a million different impartial judges, and my attorney is right that 70 per cent of these trials would result in a punishment of ten years and 30 per cent in a punishment
of zero, the ‘average’ punishment I would receive in the million trials would be seven years per trial. Going to trial is thus like being subjected to a punishment lottery that, ‘on average’, costs me seven years per trial.

Suppose the prosecutor tells me that if I plead guilty to simple robbery, he will guarantee a sentence of seven years. Then one could say that the ‘objective values’ of the plea bargain and the punishment lottery are the same: take the deal and get seven years for certain, or enter the lottery and expect to pay seven years for an ‘average’ trial. Whether I take the deal or not depends on how I feel about the risk posed by the lottery. If, for whatever reason, I am especially terrified by the thought of going to prison, I might look at the lottery pessimistically and focus on the awful prospect of ten years behind bars. Seven years for certain seems a better outcome for me now than the prospect of serving even more if I am convicted at trial. If so, I am risk averse, and will take the deal. Because the 70 per cent chance of a sentence as long as ten years seems worse to me than a certain sentence of just seven, I would even be willing to accept an even lower price for my plea, a bargained sentence a little greater than seven years, just to be relieved of the possibility of serving ten. But if prison does not hold all that much terror for me, perhaps because I have been there before and think of it as part of criminal life, or because I can pursue valuable relationships inside (or outside) the prison while I am there, I will enter the punishment lottery by refusing the bargain and going to trial. Because I look optimistically at the trial as a chance to walk free rather than pessimistically as a chance to serve ten years, I prefer the risk. To get me to relinquish my chance for an acquittal, despite the odds against it, the prosecutor will have to offer a deeper discount, a higher price, for my guilty plea, a certain sentence of less than seven years.31

It is thus not enough to depict the defendant as concerned solely with choosing the option that yields the smallest possible ‘objective’ sentence, in which case every defendant faced with this situation would accept a bargain at a day under seven years or less and refuse it at a day over seven years or more. Different defendants value a year in prison differently. A year in prison, or each year of a sentence of more than a year, imposes a cost, a quantum of distress or suffering, on defendants that differs from defendant to defendant, and, for any given defendant, from year to year. These differences produce different attitudes toward the risks of the punishment lottery. Suppose we define the variable $p$ as the likelihood, or probability, that any defendant will in fact be convicted at a full trial. This probability can vary from zero, which means that the defendant is certain to be acquitted at trial, to one, which means that he is certain to be convicted. In the example above, then, the defendant’s estimate that the chance that he will be convicted at trial is 70 per cent would mean that he estimates the probability $p$ as 0.7. Now define the variable $B$ as the sentence that the prosecutor offers the defendant in exchange for his guilty plea, and the variable $T$ as the sentence the defendant would receive were he to be convicted after a full trial. $B$ is always less than $T$, and the difference between them is the sentence discount associated with the negotiated sentence $B$. Then, as above, the defendant faces an expected punishment of $p \times T$ if he choose to exercise his right to a full trial.

If all that a defendant cared about was the severity of the punishment he would have to serve, the choice between a negotiated punishment $B$ and the prospect of a full trial with an expected punishment $p \times T$ would turn on the relative values of the two punishments: if $B$ were smaller than $p \times T$, the defendant would accept the bargain and the punishment $B$, but if $B$ were greater than $p \times T$, he would reject the bargain and take his chances at a trial. But

31 Ibid at 216–217.
because different defendants value the consequences of punishments differently, the relevant comparison is between the cost imposed on the defendant by the bargained sentence $B$ and the expected cost of going to trial and risking the sentence $T$, the defendant’s ‘average cost per trial’. We can represent the way any given defendant experiences a particular punishment $X$ mathematically by defining a cost function $C(X)$ for that defendant. This mathematical relationship, which differs from defendant to defendant, transforms the objective punishment $X$ into a number, $C$, that represents the defendant’s subjective experience of distress or suffering at being forced to endure $X$. If we assume, reasonably, that $C$ increases as $X$ increases, that is, that defendants see more punishment as worse than less, then the question actually posed by the plea offer is whether $C(B)$ is less than $p \times C(T)$. If so, the defendant will agree to plead guilty and accept the sentence $B$. If not, he will reject the agreement and enter the punishment lottery.

This formulation suggests three distinct ways that a prosecutor can induce a defendant to plead guilty, that is, make the left side of the inequality smaller, or the right side larger, for that defendant. All other things being equal, the trial alternative can be made less attractive relative to the bargain by increasing either $T$, the sentence on conviction at trial (perhaps by invoking a mandatory minimum sentence statute), or $p$, the likelihood the defendant will be convicted, or both. Or the bargain can be made more attractive relative to the trial by lowering $B$, the sentence the defendant will serve if he pleads guilty. These variables correspond to the three principal bargaining tactics employed by prosecutors in adversarial systems everywhere, with each raising the question of constitutional propriety in a different way. Most commonly, prosecutors attempt to induce guilty pleas simply by deepening the sentencing discount (‘sweetening the deal’), paying high prices for pleas by offering bargains at low values of $B$. This is the form taken by the archetypical plea bargain, so deciding on the propriety of this tactic is tantamount to deciding on the propriety of the practice itself. But even if this propriety is conceded in general, if prosecutors are so desperate for pleas that they offer extremely low values of $B$ relative to $T$, say a three-month sentence for a weapons offence for a defendant facing the possibility of ten years for aggravated robbery after a trial, the very sweetness of the deal might induce even a truly innocent defendant who thinks there is even a small chance that he will be convicted at trial to plead guilty, whether he prefers risk or is averse to it.

Alternatively, prosecutors can increase the likelihood of conviction by devoting more resources to investigation or, as is often done in cases involving more than one defendant, securing the testimony of one defendant against another by offering the defector a chance to plead guilty to a lesser offence in exchange for significantly increasing the likelihood that his co-defendant will be convicted at trial, and thus helping persuade the co-defendant to plead as well. It is precisely this situation that economists describe as the prisoner’s dilemma. Given the propriety of bargaining itself, this common tactic creates no problem of convicting the innocent. Why? Because in any adversarial system, including South Africa’s, the only authoritative institution that exists for discovering, or pronouncing, anyone guilty or not guilty is the criminal process itself. If the evidence is strong enough to convince a trial judge beyond a reasonable doubt that a defendant is guilty of a crime, he is guilty, and subject to punishment, even if he really did not do it. It is the conviction that creates the legal fact of guilt, irrespective

---

of what the defendant or anyone else may believe. Subjecting defendants to the prisoner’s dilemma thus cannot be unfair to the co-defendants who become more likely to be convicted, because finding defendants guilty or not as a legal fact, not finding authoritatively exactly who did or did not do what, is what adversary trials are about, and increasing the likelihood that a defendant will be found guilty at a fair trial is what prosecutors are supposed to do.

The third strategy, increasing $T$, is more problematic. In the United States, this is typically achieved by ‘overcharging’, piling a host of distinct charges on a defendant who has committed several offences in the course of a single criminal transaction. In *Brady v United States* (1970), in which the Supreme Court first gave its approval to plea bargaining in general and the first two tactics in particular, the Court stressed that the defendant’s participation in a plea bargain must be truly voluntary, and hinted that egregious overcharging might so ‘overbear the will of the defendant’ as to effectively coerce him into pleading guilty to lesser charges. But eight years later, in *Bordenkircher v Hayes*, the Court was forced to face the harsh realities of American plea bargaining, conducted in an environment of high crime, inadequate resources to prosecute it and, by international standards, very severe prison sentences for almost every serious offence. Hayes, who had two prior felony convictions, was charged with passing a bad cheque for less than $90, an offence punishable by imprisonment of two to ten years. The prosecutor offered Hayes a five-year sentence for his guilty plea, and to pressure him to accept, told him that if Hayes were convicted at trial on the cheque charge, he would invoke the state’s habitual offender act, a mandatory sentencing statute that provided for an automatic sentence of life imprisonment after a third felony conviction. Hayes, fully aware of this threat, nonetheless refused the bargain, was convicted at trial and, when the prosecutor invoked the habitual offender act as threatened, duly sentenced to life imprisonment.

But because it had to in order to preserve the plea bargaining system and, with it, the entire American criminal process, the Court held that the threat and eventual invocation of the habitual offender act was an acceptable prosecutorial bargaining tactic. If an injustice was done to Hayes in this case, it suggested, it was not the prosecutor’s threat that did it, but the fact that the state’s habitual offender act, which was not challenged in the case and which all sides agreed Hayes had violated, permitted a life sentence in these circumstances. If plea bargaining is to work, defendants must know that the sentencing discount is real. The only way any defendant can be induced to plead guilty in exchange for a proffered discount is to turn those, like Hayes, who do insist on a trial into admonitory examples for the rest, by following through on the threat to sentence them more harshly when they are convicted than they would have been had they pleaded guilty. Prosecutors who need to induce pleas, the Court reluctantly recognised, must be permitted to credibly threaten recalcitrant defendants in this way and, as in *Bordenkircher*, to use existing statutes that defendants have in fact violated to do so:

> Indeed, acceptance of the basic legitimacy of plea bargaining necessarily implies rejection of any notion that a guilty plea is involuntary in the constitutional sense simply because it is the end result of the bargaining process . . . . [By] tolerating and encouraging the negotiation of pleas this Court has necessarily accepted as constitutionally legitimate the simple reality that the prosecutor’s interest at the bargaining table is to persuade the defendant to forego his right to plead not guilty.

33 Cf *Herrera v Collins* 560 US 390 (1993)(Petitioner convicted of capital murder is not entitled to *habeas corpus* relief simply because, years after his conviction, new evidence tending to prove his innocence is discovered).

34 *Brady v United States* 397 US 742, 750 (1970); Adelstein (note 32 above) at 823–827.

35 *Bordenkircher v Hayes* 434 US 357 (1978); Adelstein (note 25 above) at 227–229.

36 *Bordenkircher v Hayes* (note 35 above) at 361–364.
C Information and uncertainty

Defendants, and only they, know their own cost functions and the attitudes toward punishment these reflect. But in order to decide how to plead, a defendant also needs information about three variables, $B$, $p$ and $T$. These are ‘objective’ in the sense that they can be expressed numerically, but the information the defendant has about them may be imperfect, forcing him to make largely subjective estimates on the basis of the information he has about what the true values of the objective variables are. In many situations, defendants do have about as much of the information as it is possible for them to have. If the prosecutor is clear as to what value of $B$ he is offering and the defendant knows both that the prosecutor will not renege on the offer in court and that the sentencing judge will respect the agreement and impose the sentence $B$; if the defendant can rely on a well-informed guess (all that is possible in any case), formed either on the basis of his own experience or the advice of counsel, as to the likelihood that he will be convicted; and if he can be confident as to what sentence he would receive if he were convicted at trial, then the decision the defendant makes as to plea can be said to be fully informed, and absent coercion – a question we will soon revisit – a reflection of his best interests as he sees them in a painful predicament. But if not, apart from whatever normative questions of fairness or constitutionality might arise from allowing defendants to plead guilty in the absence of reliable information about $B$, $p$ and $T$, where defendants must in fact decide whether to plead guilty in substantial ignorance of these variables, the plea bargaining system itself is likely to perform poorly in the role the criminal process has assigned to it.

How do defendants acquire this information and assess its reliability? Some defendants, as suggested above, will be able to rely on personal experience. Having encountered the criminal process before, they may be aware of the sentencing practices of trial judges, have a good sense of what the likelihood is that they will be convicted on the available evidence, and perhaps even be able to judge whether the prosecutor is offering them the best deal they can get or if a bit of resistance might result in a better one. But whether they are experienced or not, personal knowledge, from whatever source it might be gleaned, and the representations and promises of the prosecutor are all the information that most defendants in CPA s 112 proceedings have. For any defendant, the best source of reliable information during the bargaining process about all the relevant variables, and the person best positioned to wrangle the best bargain possible from the prosecutor, is a competent defence attorney, and counsel is a luxury most CPA s 112 defendants do not enjoy. Without an attorney, defendants, typically at a moment filled with stress and anxiety, must evaluate not just the likelihood of conviction at trial and the sentence that will be imposed thereafter, but the precise terms and legal consequences of the proffered discount and its value relative to the trial sentence, by themselves. Compared to competently represented defendants, these circumstances put them at a substantial disadvantage in identifying and concluding agreements that actually serve their interests. Should they err in estimating the key variables and agree to a bargain they come to regret, once entered and accepted by the judge, the deal becomes final, just as it does for represented defendants. In *State v Armugga*, Msimang J noted as follows in refusing a request to reopen a s 105A bargain after the defendant learned he had agreed to a less favourable deal than he thought he had:

> In the process of bargaining, numerous assumptions are made and mistakes are bound to happen. Provided that a party is found to have acted freely and voluntarily, in his or her sound and sober senses and without having been unduly influenced when conducting a plea bargaining
agreement, the fact that the assumptions turn out to be false, does not entitle such a party to resile from the agreement.37

In addition to an attorney to advise them about \( T \) and \( p \), s 105A defendants also have the advantage of having the agreement in all its particulars reduced to writing, so that the true value of the prosecutor’s offer \( B \), the magnitude of the discount it represents and the full legal consequences it entails, are made as clear as possible to them. If, having agreed to plead guilty in return for the sentence specified in the agreement, the defendant is surprised at the judge’s refusal to impose it in court, he may withdraw the plea and demand a full trial. His s 112 counterpart, in contrast, represented or not, will have no opportunity to withdraw his plea even if, having agreed to an informal deal with the prosecutor and pleaded guilty in reliance on it, he finds the judge unwilling to honour the agreement and determined to impose a sentence greater than \( B \). Even where the information the defendant needs exists in reliable form, where, that is, prosecutors offer clear terms with no intention of reneging and judges are prepared to sentence as the agreement specifies, the procedures of s 105A are clearly superior to the practices of s 112 in providing defendants with the information that they need to ensure that their decisions represent, as far as practicable in a challenging environment, their own interests as they see them. Where information exists but is unreliable, not only are the difficulties compounded for individual defendants already faced with tough decisions – whether they have legal representation or not – but serious questions are raised about the dysfunction and fairness of the plea bargaining system itself.

Because every criminal case is different, and because at the moment defendants must plead, the trial lies in an unknowable future, it is hard to see how an experienced defence attorney’s estimate of \( p \), the likelihood of a particular defendant’s conviction, can be improved by institutional design. The punishments in the statute books themselves, supplemented perhaps by informal knowledge of the sentencing habits of particular judges, are generally sufficient to give defendants a useful predictor of \( T \). But misinformation, not simply error, about the variable \( B \), the precise value of the sentencing discount the defendant is being offered for his guilty plea, is often the result of specific actions, intentional and unintentional, by prosecutors and judges that can effectively be policed and remedied by statute and judicial oversight. If left unaddressed, these errors and omissions will erode the bargaining system’s ability to identify and facilitate mutually beneficial bargains.

Recall that, given \( B \), \( p \) and \( T \), a defendant will plead guilty if the subjective cost to him of the bargained sentence is less than the expected cost of going to trial, that is, \( C(B) < p \times C(T) \). Suppose, however, the possibility exists either that the prosecutor will renege on his offer when it comes time for him to perform, say by making a different sentencing recommendation to the court than that provided in the agreement, or that the judge, who ultimately remains in possession of the sentencing power, will not abide by the terms of the agreement and sentence the defendant more harshly than the agreement demands. In either case, if the defendant thinks there is a chance that, having been promised a sentence \( B \), he will actually receive a more severe punishment and not be able to extract himself from the bargain, he will have to replace \( B \) in his calculations with the higher sentence \( (B + h) \), where \( h \) represents the expected increase in the sentence created by the possibility that the prosecutor will renege on his promise or the judge will impose a sentence greater than \( B \). If there are enough broken promises or unexpected judicial nullifications of bargains so that \( h \) is significant, plea bargains that would have served

37 State v Armugga 2005 (2) SACR 259, 265c–d (N).
not just the interests of the immediate parties to it but the larger interest of the criminal process in diverting trials to less costly procedures as well, and that would have been made without this uncertainty, won’t be made because of it. This will occur whenever the cost to the defendant of a bargain at $B$ is smaller than the expected cost of the trial but the cost of a bargain at $(B + h)$ is greater than the expected cost of the trial, that is, when $C(B) < p \times C(T) < C(B + h)$. The existence of even a few unkept bargains will be enough to surround every subsequent negotiation with an uncertainty that keeps at least some otherwise mutually beneficial bargains from being concluded. Expensive trials that no one wants will have to be mounted, and the already burdensome costs of criminal justice will rise still further. Just as bad money drives good money from circulation, bad bargains will drive out good ones, to everyone’s detriment.  

In 1971, in *Santobello v New York*, the United States Supreme Court addressed the first of these sources of misinformation, the failure of the prosecutor to keep the promises made in the agreement. Santobello, charged with two gambling offences, agreed to plead guilty to one in return for the prosecutor’s promise not to recommend a prison sentence. But between the time the agreement was concluded and Santobello was to be sentenced, a new prosecutor had taken over the case and, unaware of the promise his predecessor had made, recommended that Santobello be sentenced to a year in prison on his guilty plea, and the judge complied. Santobello protested, but was not allowed to withdraw his plea. On appeal, the Court, speaking through Burger CJ, found for Santobello. It began by explicitly acknowledging the inability of the American criminal process to survive without plea bargaining, something the Court in *Brady* had refused to do the previous year, grounding its approval of bargaining instead on the ‘mutuality of advantage’ it offers to both prosecution and defence. Plea bargaining, the Santobello Court now said, is ‘an essential component of the administration of justice. Properly administered, it is to be encouraged’. But proper administration requires ‘fairness in securing agreement’, and in ensuring it the Court applied a basic principle of ordinary contract law to hold that, despite the fact that the breach of the agreement was inadvertent and the one-year sentence itself not unjustified, Santobello must receive specific performance of the agreement or be permitted to withdraw his plea: ‘[W]hen a plea rests in any significant degree on a promise or agreement of the prosecutor, so that it can be said to be part of the inducement or consideration, such promise must be fulfilled’.  

For decades before 1970, the Supreme Court turned a blind eye to the unpleasant business of plea bargaining. Only after its hand was forced by some careless language in a 1968 opinion not involving a plea bargain at all did the Court discuss plea bargains in detail and start regulating them in *Brady* and *Santobello*. In South Africa too, plea bargaining under s 112, increasingly acknowledged by scholars and practitioners, remained in the judicial shadows until 1999, when *North Western Dense Concrete CC v Director of Public Prosecutions* was decided by the High Court on facts much like those in *Santobello*. The applicant and its employee were

38 Adelstein (note 32 above) at 814–816.
40 Ibid at 260. See also *Brady v United States* (note 34 above) at 752.
41 *Santobello v New York* (note 39 above) at 260, 261, 262. Note that reaching the ‘fair’ outcome in this case also had the effect of ‘making plea bargaining work better’. This confluence of fairness and efficient operation in the bargaining system is discussed further in R Adelstein ‘Institutional Function and Evolution in the Criminal Process’ (1981) 76 Northwestern University Law Review 1, 71–79.
43 Compare Bekker (note 3 above) at 218; Steyn (note 1 above) at 206; Kerscher (note 16 above) at 30.
charged with culpable homicide, and the prosecutor agreed to drop the charge against the applicant if the employee pleaded guilty, which he did in a s 112 proceeding. But after another party had petitioned the respondent for a nolle prosequi in the case, the DPP reconsidered its position and reinstituted the charge against the applicant, in violation of the earlier plea agreement.\(^\text{44}\)

Uijs J’s opinion closely parallels those of the Supreme Court in *Brady* and *Santobello*. He began by acknowledging the long history of informal plea bargaining under s 112 in South Africa, a provision ‘tailor-made’ for plea bargains, and lamented the judiciary’s unwillingness to confront it openly.\(^\text{45}\) The court conceded, as the Supreme Court had in *Santobello*, that bargaining under s 112 was an integral component of South African criminal justice, one so entrenched that eliminating it was likely to cause the breakdown of the entire criminal justice system\(^\text{46}\) and, like the *Santobello* Court, appealed to ordinary contract law to draw the necessary conclusion. Because it would be unfair to allow the respondent to enjoy the benefits of the agreement without meeting the obligations it created for him, the court granted a permanent stay of prosecution to the applicants, effectively ordering specific performance of the contract. It held that a guilty plea under s 112, if accepted by the prosecutor, binds both the prosecutor and the court on the facts stipulated in the agreement, and requires the prosecutor to honour the terms on which it has been reached.\(^\text{47}\) As in *Santobello*, requiring prosecutors to fulfil the agreements on which defendants have relied in pleading guilty serves not just the interest of fairness to defendants, but goes far toward preserving the system’s ability to facilitate the completion of agreements that represent the defendant’s informed judgment of his own interests in the case.

The second source of uncertainty for defendants is whether the sentencing judge will abide by the agreement and impose the sentence it calls for. If the defendant fears that the sentence he receives will be greater than he bargained for, he will have to add the increment \(h\) to his assessment of the bargain’s terms, reproducing the problem caused by breach of the agreement by the prosecutor. But the independence of sentencing judges makes securing their cooperation in the completion of plea bargains by sentencing in accord with them a delicate matter. One approach, adopted by the American federal courts for plea agreements couched in carefully specified language, is to explicitly withdraw the sentencing power from the judge in these cases and require the bargained sentence to be imposed.\(^\text{48}\) Another, more implicit approach employed successfully by most of the states is not to direct the judge to impose the sentence specified in the agreement but to rely instead on administrative judges, whose interest in relieving the enormous pressure on the criminal process is closely aligned with the prosecutor’s, to ensure that sentencing judges do what they must to keep the system afloat.

South Africa’s approach differs across the two tracks to agreement. Because judges retain the full authority to impose any lawful sentence after a guilty plea in a s 112 proceeding, informal bargaining in that context can only be charge bargaining, with the parties’ agreement on the facts and the charge the principal constraint on the judge’s choice of sentence within the statutory range. If a defendant believes there is a significant likelihood that the judge will

\(^\text{44}\) *North Western Dense Concrete CC* (note 6 above) at 671–672.

\(^\text{45}\) Ibid at 672–677.

\(^\text{46}\) Ibid at 683f, 676f and 678c.

\(^\text{47}\) Ibid at 670f-g, 677b. See also South African Law Commission ‘Sentence Agreements’ (note 12 above) at 3.16.

not impose the sentence he and the prosecutor hope she will, this uncertainty will distort his choice as we have discussed, and his reluctance to bargain will only be increased by his inability to withdraw his plea should he discover after he has entered it that the sentence is greater than he bargained for. But in severely overburdened criminal courts, as in the United States, to the extent that trial judges respond to the pressures that heavy caseloads put on the judiciary and the prosecution alike, they too may tacitly agree to honour the bargains put before them and keep the cases moving along.

Section 105A seems to recognise this problem, and takes steps to alleviate it. The written agreement that manifests the formal bargain must not only ‘state fully the terms of the agreement, the substantial facts of the matter, all other facts relevant to the sentence agreement and any admissions made by the accused’, it must also propose a specific punishment that, in the view of the parties, would constitute ‘a just sentence to be imposed by the court’.\(^\text{49}\) First, the judge questions the defendant to determine that he understands the proceedings and the agreement. Second, she establishes a factual basis for the plea.\(^\text{50}\) Third, she then considers the sentence agreed upon by the parties. If she finds the agreement ‘just’, she must then convict the defendant and sentence as the agreement requires. If she finds the agreement ‘unjust’, she must inform the parties of what alternative sentence she would consider just, and give them an opportunity to revise the agreement on the basis of this sentence and resubmit it to her for sentencing. If either side declines to accept these terms, the plea is withdrawn and the defendant is entitled to a new trial \textit{de novo} before a different judge.\(^\text{51}\)

Steyn points out that the statute’s use of the word \textit{just} to describe an acceptable sentence differs from the more traditional usage in this context, \textit{appropriate}, and suggests that this shift is meant to signal to judges that they should approach the sentencing of defendants in s 105A proceedings differently than they do the sentencing of defendants convicted after a full trial.\(^\text{52}\) Two important cases support this view. In \textit{State v Sassin}, the High Court considered the sentence imposed in a s 105A bargain.\(^\text{53}\) Majiedt J refused to interpret ‘just’ to imply that the sentencing judge is required to agree with the proposed sentence, despite his suggestion that, had he been the judge in the case, he would have imposed a longer sentence. Why? Because the trial judge took proper account of the ‘triad’ of traditional determinants of appropriate sentencing enunciated in \textit{State v Zinn}:\(^\text{54}\) (1) the seriousness of the offence; (2) the interests of society; and (3) the personal circumstances of defendant.\(^\text{55}\) In this way, Watney notes, \textit{Sassin} encourages judges to take account of ‘the same factors a sentencing court would normally consider but with the proviso that any sentence that could possibly be considered appropriate under the circumstances would suffice’.\(^\text{56}\)

Two years after \textit{Sassin}, in \textit{State v Esterhuizen},\(^\text{57}\) the High Court revisited the question of ‘justness’ in s 105A matters. The sentencing judge’s role, wrote Els J, was not merely to consider the ‘well-known triad’, but also to take ‘a broad overview of the facts admitted . . . together

\(^{50}\) Criminal Procedure Act 51 of 1977, ss 105A(5–6).
\(^{51}\) Criminal Procedure Act 51 of 1977, ss 105A(7–9).
\(^{52}\) Steyn (note 1 above) at 213–215.
\(^{53}\) \textit{State v Sassin} 2003 (4) All SA 506 (NC).
\(^{54}\) \textit{State v Zinn} 1969 (2) SA 537 (A).
\(^{55}\) \textit{State v Sassin} (note 53 above) at 15.5–15.8, 18.1.
\(^{56}\) Watney (note 16 above) at 227.
\(^{57}\) \textit{State v Esterhuizen} 2005 (1) SACR 490 (T).
with the proposed sentence to be imposed, all with a view to establishing whether the sentence agreed upon and its effective content bear an adequate enough relationship to the crimes committed – taking into account all of the agreed facts, both aggravating and mitigating, so that it can be said that justice has been served.’ 58

Like Majiedt J, Els J noted that, as a sentencing judge, he ‘would probably have imposed a much heavier sentence under the circumstances’, but ‘it must be so that the court, in considering the “justness” or “unjustness” of a sentence agreement cannot simply decide for itself in vacuo what sentence it would have imposed for crimes to which the accused is pleading guilty’. 59 In a case that is difficult to prove and the likelihood of conviction correspondingly low, but where the prosecutor is convinced of the defendant’s guilt, a plea bargain may well result in a smaller sentence than might be imposed after a successful trial. But this alone would not render the sentence unjust:

In return for the concession of a plea of guilty to a charge difficult to prove, it must be so that the Legislature has envisaged that the bargaining mechanism would bring home a result which satisfies the interests of justice. These would be that where a crime has been committed a conviction has been achieved. The price may be that the sentence which would normally flow from the commission of such a crime is lower than might otherwise have been imposed. This does not mean that justice has not been achieved. 60

The High Court seems to be treading a very narrow path. At the same time the High Court is preserving nominal judicial discretion in sentencing and attention to the traditional factors of sentencing even in the presence of a s 105A agreement, it is encouraging judges to agree to negotiated sentences that may be far less severe than those that would result after a conviction at trial, even (or perhaps especially) in cases where imperfect evidence makes conviction at trial hard to achieve. It seems to be inviting, or asking, as subtly and delicately as it can, sentencing judges to swallow hard and impose whatever sentence the parties may specify in a plea agreement, so long as it bears a reasonable relationship to the crime the defendant has admitted to committing. This compromise, as Judge Els put it, is the ‘price’ that must be paid for the system’s commitment to plea bargaining.

III IS PLEA BARGAINING CONSTITUTIONAL?

A Fairness and coercion

There are good reasons to think that plea bargaining is a bad thing. For one, it produces inaccurate outcomes, not once in a while, or even inevitably, but systematically. Allowing offenders to be punished for crimes less serious than what they have actually done offers prosecutors several advantages. As Els J suggests, it lets them successfully conclude a larger portion of cases with convictions, and deploy their scarce human and material resources to inflict some punishment, even a small amount, on as many offenders as possible, all of which is surely in the interests of the criminal process, if not necessarily justice. But when huge caseloads put their resources under great stress, so the price in punishment prosecutors are willing to pay for a guilty plea is very high, defendants routinely plead guilty to crimes that bear only

58 Ibid at 495a–b.
59 Ibid.
60 Ibid at 494c, 494b, 494h. Cf Watney (note 16 above) at 228–229. But see Bennun (note 17 above) at 32 (Criticising this view as ‘a serious misconception of the nature of a criminal trial’.)
faint resemblance to, and suffer punishments far smaller than the law prescribes for, what they have actually done. And because the essence of the bargain is the sentencing discount, the bias is always in the same direction – every bargain results in a sentence smaller than what the defendant could expect after a trial on an accurate charge. When discounts are deep, punishments vastly understate the moral weight of the offences that were actually committed, so that when this systematic effect is added to the uncertainty already attached to the law’s inability to capture and convict every offender, the expected punishments that prospective offenders actually face for their crimes become very much lower than the moral harm done by their crimes. To the extent that punishment deters crime, this result means that many more crimes will be committed than would be the case than were the punishments authorised by statute for the crime actually committed imposed in every case.

A second objection is that plea bargains remove the doing of criminal justice from public view. Charges and sentences are negotiated informally under s 112, often hastily and under pressure, by prosecutors and defendants, or by prosecutors and defence attorneys, who must then negotiate with their clients, far from public view. Even in a s 105A bargain, the only evidence ever brought to light is what the parties choose to incorporate into the written agreement and the defendant admits in colloquy with the judge, and the only members of the public permitted even indirect access to the proceedings are the victims, whose views prosecutors are required to solicit. In this way, Bennun argues, plea bargaining subverts adversarial norms by supplementing the prosecutor’s already substantial charging discretion with the further power to determine the legal fact of guilt or innocence, traditionally the province of the trial court, and denies the public its interest, distinct from that of the defendant, in the defendant’s right under s 35(3)(c) of the Constitution to ‘a public trial before an ordinary court’. In the United States, where public defenders do a great deal of the bargaining for impecunious defendants, defendants may be pressured to accept agreements by their attorneys, hoping to maintain cooperative relationships with prosecutors at the expense of the particular client whose fate they are deciding. It is not a pretty sight, even if we could see it.

‘Even if we could see it’ – sub rosa justice is a third objection. Criminal trials have a kind of majesty, as the full power of the state is deployed against a lone defendant purposely equipped by the law with rights to resist it effectively. Conducted with due regard to these rights, they help make ordinary people into citizens by inviting them to observe how the government treats suspected offenders, what standards it applies to them and itself, and when and how it makes mistakes. Staying its hand against defendants until fair procedures convict them shows the modern state at its best, elevated in its citizens’ eyes by the respect it shows them and the law. Plea bargains, as Bennun and many others have pointed out, possess none of this appeal or obvious normative legitimacy. Compared to the moral weightiness of a full trial, they seem shabby, even dirty, like dispensing rough-and-ready justice in a dusty bazaar. Judged against the normative and constitutional standards manifested in a modern criminal trial, the undignified, unlovely haggling of plea bargaining is hard to defend on any ground beyond necessity. Not even its defenders like it much.

But to say that plea bargains are ugly, or even that they produce systematic inaccuracies, is not necessarily to say that they are unconstitutional. The fate of plea bargaining in the American federal courts is instructive in this regard. As early as 1908, in Twining v New Jersey, the Supreme Court had recognised that

---

61 Bennun (note 17 above) at 32–34, 37–43.
some of the personal rights safeguarded by the first eight Amendments against National action may also be safeguarded against state action, because a denial of them would be a denial of due process of law. If this is so, it is not because those rights are enumerated in the first eight Amendments, but because they are of such a nature that they are included in the conception of due process [guaranteed by the fourteenth amendment].  

After World War II, the Supreme Court applied the criterion of ‘fundamental fairness’ to various aspects of state criminal procedure as they came under constitutional challenge. The Court acted on Twining’s promise by gradually ‘incorporating’ almost the full range of rights guaranteed to suspects and defendants by the first eight amendments into the due process clause of the fourteenth amendment, thus making them binding on the states as well as the federal government. But until Brady was decided in 1970, as negotiated pleas increasingly colonised both the state and federal criminal courts and the regime of plea bargaining came under increasing constitutional pressure in the lower federal courts, the Supreme Court refused to acknowledge the existence of bargaining or test its constitutionality against the fifth amendment right against self-incrimination and the sixth amendment right to a criminal trial before a jury. It regulated the taking of guilty pleas only by observing that they represented a waiver of the defendant’s right to trial and a conviction on the basis of the plea alone, and requiring that they be made ‘knowingly’, in full awareness of their consequences for the defendant, and ‘voluntarily’, without physical or mental coercion that might deny a defendant freedom of choice in waiving the right to trial.

The problem of knowledge attracted little judicial attention. It could effectively be addressed by careful design of the rules that govern the bargaining process itself, as indeed it has been, legislatively by s 105A in South Africa and in the United States by judges and judicial administrators after 1970. Instead, the voluntariness of guilty pleas became the principal field on which the constitutionality of negotiated pleas was contested. In one sense, absent physical duress, it is hard to see how a defendant who knowingly trades his right to trial for the advantage he perceives in the lower sentence the prosecutor is offering for his guilty plea might be acting involuntarily. The fact that the defendant must choose the lesser of two evils, pain or the chance of more pain or none, does not make his choice to plead guilty involuntary. But the close relation between the due process clause of the fourteenth amendment and the definition of constitutionally acceptable criminal procedure in American law means that the construction of every right guaranteed by the fourth, fifth, sixth or eighth amendments is strongly inflected by concerns of ‘fundamental fairness’. Accordingly, the Supreme Court has taken a subtler view of voluntariness and coercion. As Frankfurter J put it in 1948, holding that a station-house confession by a fifteen-year-old boy was not a voluntary act:

It would disregard standards that we cherish . . . to hold that a confession is ‘voluntary’ simply because the confession is the product of a sentient choice. . . . [C]onduct devoid of physical

63 The scholarly literature on incorporation of the first eight amendments into the due process clause of the fourteenth is vast. See, eg, AR Amar ‘The Bill of Rights and the Fourteenth Amendment’ (1992) 101 Yale Law Journal 1193 (Amar offers a perceptive overview of the literature).
64 The fifth amendment right against self-incrimination was incorporated into the fourteenth amendment and applied to the states in Malloy v Hogan 378 US 1 (1964), as was the sixth amendment right to a jury trial in Duncan v Louisiana 391 US 145 (1968).
65 Adelstein (note 32 above) at 816–820; Alschuler (note 3 above) at 33–38.
pressure but not leaving a free exercise of choice is the product of duress as much so as choice reflecting physical constraint.\footnote{Haley v Ohio 332 US 596, 606 (1948).}

In this pre-1970 tradition, the federal appellate courts confronted a range of issues in considering the voluntariness of guilty pleas. Their primary concern was the danger that innocent defendants would falsely condemn themselves by accepting the offer of a sharply discounted sentence and pleading guilty to offences that they did not in fact commit. The problem of erroneous convictions thus became intertwined with the voluntariness of the plea, and the justiciable question that emerged was whether a given bargaining tactic, or the plea bargaining process itself, created a substantial risk that innocent defendants would waive their right to trial and convict themselves by falsely pleading guilty to some offence.

How might plea bargains be seen as coercive in this sense? The model of part II suggests a useful approach. Recall our earlier example: I am accused of aggravated robbery, which I may or may not have done. The probability that I’ll be convicted at trial is 70 per cent, and if I am, I’ll be sentenced to ten years in prison, so the expected trial sentence is \(0.7 \times 10 = 7\) years. As we saw, if the prosecutor offers me a sentence of seven years in exchange for my guilty plea, the ‘objective values’ of the bargain and the punishment lottery are the same: accept the bargain and get seven years for certain, or enter the lottery and pay seven years for an ‘average’ trial. Whether or not I take the deal depends on my attitude toward the risk of the lottery: if I am a risk averter, I will take the deal, but if I am a gambler, I will take the risk of trial. But in neither case am I forced to ‘pay a price’ for exercising my right to trial. As far as anyone knows before the trial, the ‘objective’ punishments in the two alternatives are the same, and whether I take the deal or not depends on my preferences and the attitudes toward risk they imply. It is only when the price the prosecutor is willing to pay for the plea, the extent of the sentencing discount, is greater than the objective value of the plea that a price is put on exercising the right to trial. It is the magnitude of this price that raises the question of coercion.

Suppose the prosecutor offers a sentence of five years. Then the price he is willing to pay for the plea, five years less than the trial sentence, is greater than the three-year discount needed to make the definite punishment of the bargain and the expected punishment in the lottery have equivalent objective value. Accepting the bargain means five years for me, when the expected sentence after trial is seven, so exercising my right to trial means paying an objective price of two years. Whether I surrender my right and accept the bargain or pay the two-year price for exercising it depends, again, on my preferences. But in either case, refusing the deal means paying an objective price of two years for exercising the right to a trial.

Now, as prosecutors in large American cities must do every day, turn up the heat. Suppose a hard-pressed prosecutor offers me nine months for my plea. This may be an offer I cannot refuse, a bargain that, given my predicament, is much too good to pass up no matter how I feel about the risk of trial. In light of the possible outcomes, the price I would now have to pay to exercise my right to trial is very high. Even if I know I am not guilty, and even if I am a gambler and time in prison holds less terror for me than it does for risk averters, if there is a 70 per cent chance that I will have to serve ten years if I go to trial and I can avoid this risk by serving just nine months, I am very likely to accept the deal and plead guilty. The overwhelming need of American prosecutors to secure guilty pleas has, for at least a hundred years, meant that terms
like this are offered to most American defendants, burdening their choice as to plea by making exercise of their right to trial a very costly proposition.67

This analysis came to the fore in federal appellate courts throughout the 1960s. In 1969, Bazelon J proposed that courts and legislatures had an obligation to determine when the discounts prosecutors were offering created a significant risk that innocent defendants would plead guilty in response, and to develop guidelines to help prosecutors avoid crossing this threshold of impermissibility.68 England ultimately adopted this approach. In 1994, it largely obviated the need for explicit bargaining by establishing a fixed price for guilty pleas in every case: a one-third reduction in the statutorily prescribed sentence if the plea is made at the first reasonable opportunity; a one-quarter reduction once the trial date has been set; and a one-tenth reduction if the plea comes after the trial begins.69 However, in Brady and Bordenkircher, the US Supreme Court created no such guidelines and put few limits on the discounts prosecutors may offer to secure guilty pleas, even where these are so great as to impose a significant price on a defendant’s exercise of the right to trial. And in Santobello, the Court made explicit that the reason it was tolerating such a regime was that the alternative, a full criminal trial for every defendant, was far beyond the material means of the criminal process. There was, it acknowledged, no plausible alternative to plea bargaining but chaos.

B Rights and limitations

This history illuminates two important aspects of the question of plea bargaining’s constitutionality in South Africa that distinguish it from the American case. The first is what Snyckers and le Roux70 call the due process wall, a conceptual barrier erected by the Constitutional Court to analytically separate the ‘residual liberty rights’ of s 12 of the Constitution, and in particular the right of every person ‘not to be deprived of freedom arbitrarily or without just cause’ guaranteed by s 12(1)(a), potentially broad, due process-like rights that might well be informed by judicial notions of justice or fairness, from the specific rights of arrested, detained and accused persons in the criminal process enumerated in s 35. Where the American due process clause has infused the definition of fifth and sixth amendment rights with considerations of fundamental fairness, the wall is intended to prevent the residual rights of s 12(1) from ‘seeping’ into the definition of the enumerated rights of suspects and defendants and expanding them beyond what the language of s 35 itself, presumably in its traditional or common law usage, requires.

The prosecutor’s offer of nine months in our example illustrates the effect this interpretive constraint might have in the context of plea bargaining, an alternative to the full adversarial trial as a means of convicting defendants. In the United States, the seepage of due process rights into the definition of voluntary guilty pleas might lead a judge, like Frankfurter J, to

68 Scott v United States 419 F2d 264 (DC Circuit 1969).
look beyond the defendant’s ostensible consent to the bargain to conclude that the offer put too great a price on the right to trial and was thus unfairly coercive. But a South African court, forced by the wall to eschew considerations of this sort and consider the claim solely in terms of an accused’s rights to be presumed innocent under s 35(3)(h) and not to be compelled to give self-incriminating evidence under s 35(3)(j), rights themselves derived from the common law, might take a narrower version of coercion and sustain the bargain on the basis of the defendant’s informed acceptance of the offer. Indeed, Snyckers and le Roux read the series of Constitutional Court decisions that erected the due process wall itself, particularly *Nel v Le Roux NO & Others*,71 as seeking to narrow the rights enjoyed by suspects and defendants under both s 12 and s 35 of the Constitution and establishing a striking principle with clear relevance to a general regime of plea bargaining:

The sort of ‘trial’ one is entitled to under FC s 12(1)(b) [the right not to be detained without trial] differs from the sort of trial one is entitled to under FC s 35(3). The former lays down less rigorous requirements, or requirements less generous to the individual concerned, than the latter. . . . The more arbitrary or informal the proceedings in question, ie the less closely they resemble a criminal trial, the less claim the imprisoned individual has to ‘full’ fair trial rights under FC s 35. If the state were to embark upon a general practice of instituting summary proceedings which bear little resemblance to criminal trials as its main process to incarcerate those suspected of crimes, the reasoning in *Nel* would hold that the new class of criminal ‘accused’ would not be entitled to the right to a fair trial with trimmings.72

The second point is suggested by the Supreme Court’s tortured resolution of the plea bargaining question. Though the appellate courts, led by Judge Bazelon, had framed the question for the Supreme Court in terms of the price plea bargains might place on the exercise of the right to trial, the Court in *Brady* declined their invitation to examine the potentially coercive effects of large sentencing discounts and, with one eye clearly fixed on preserving what the justices knew to be an indispensable part of the American criminal process, swept them away with an appeal to the ‘mutuality of advantage’ offered by plea bargains and a heroic assumption about the state of mind of defendants who accept them:

Of course, that the prevalence of guilty pleas is explainable does not necessarily validate those pleas or the system which produces them. But we cannot hold that it is unconstitutional for the State to extend a benefit to a defendant who in turn extends a substantial benefit to the State and who demonstrates by his plea that he is ready and willing to admit his crime and to enter the correctional system in a frame of mind that affords hope for success in rehabilitation over a shorter period of time than might otherwise be necessary.73

Not until the following year, in *Santobello*, did the Court admit the practical necessity of negotiated pleas, and of adjusting the contours of the fifth and sixth amendments to it, even as it could scarcely bring itself to call the practice by its name:

The disposition of criminal charges by agreement between the prosecutor and the accused, sometimes loosely called ‘plea bargaining,’ is an essential component of the administration of justice. Properly administered, it is to be encouraged. If every criminal charge were subjected to a full-scale trial, the States and the Federal Government would need to multiply by many times the numbers of judges and court facilities.74

---

71 1996 (3) SA 562 (CC), 1996 (1) SACR 572 (CC).
72 Snyckers & le Roux (note 70 above) at 51 13–14 (emphasis added).
73 *Brady v United States* (note 34 above) at 752–753.
74 *Santobello v United States* (note 39 above) at 260.
In effect, *Brady* and *Santobello* apply a limitations analysis to the sixth amendment right to trial, in the absence of an explicit limitations clause in the American Constitution. Because it was imperative to the continued operation of the criminal process itself that the right to trial be limited by making its exercise costly (sometimes, as in *Bordenkircher*, very costly) to defendants, the Court concluded that it could not find the practice unconstitutional. This conflation in a single constitutional analysis of two distinct questions – how far a given right extends and when it might be infringed, and what practical or policy considerations might call for defining the right in one way or another – invites what Woolman and Botha call ‘the worst kind of analytical confusion’.  

But the South African Constitution *does* have a limitations clause, s 36. It requires analysis of the constitutionality of plea bargaining to proceed in two stages. First, there must be an initial determination of the content of a given right and whether it has been infringed by some specific bargaining tactic or by the plea bargaining system in general, without reference to practical considerations. Then, should the right indeed be found to have been infringed, a second, subsequent inquiry into whether the infringement can be deemed reasonable and justifiable in terms of s 36 must take place.

In light of all this, how might the constitutionality of plea bargaining under s 35 of the Constitution be tested? A brief sketch of two related claims that might be raised against both the informal bargaining system that has grown around ss 112 and 113 of the CPA and the formal bargains governed by the Act’s s 105A may be suggestive. Both focus on the adequacy of the process by which guilt is adjudicated in the regime of plea bargaining. As Bennun points out, prosecutors largely determine the ultimate outcomes of bargaining, on the basis of factors that often have little to do with finding the result in each case that most closely conforms to the governing substantive law. Before the decision to prosecute is made, they evaluate whatever evidence exists against defendants and dismiss charges against those whose guilt they might not doubt but cannot prove. Where the evidence is strong enough to proceed, they select the charges for strategic as well as legal reasons, and adjust their offers to reflect the likelihood that defendants can be convicted at all. Once they decide what discount to offer for the defendant’s plea, the defendant must accept, which may require negotiation and a further concession. Whether the deal is sealed informally by oral promises under s 112 or recorded as an enforceable agreement under s 105A, once it has been made, and before the case ever appears before a trial court, it is the parties, and primarily the prosecutor, who will have effectively made the decision as to what crimes the defendant will be found to have committed, and what evidence has or has not been sufficient to support that conviction, with little or no independent review of the decisions or the evidence. The outcome of this complex set of behind-closed-door calculations, determined in large part by the prosecutor, is then presented to a judge for ratification as ‘just’.

The presumption of innocence and its sibling, the prosecutor’s burden to prove guilt beyond a reasonable doubt, are nowhere to be seen. Section 35(3)(h) of the Constitution grants every accused the right ‘to be presumed innocent’. Snyckers and le Roux interpret this provision as ‘a right to expect the state to bear the full burden of proving the case and therefore not to be allowed to compel assistance from the accused’. The typical plea bargain in a crowded jurisdiction is unlikely to meet such a standard. The admissible evidence in the case is never

---


76  Snyckers & le Roux (note 70 above) at 51–144.
put before an independent fact-finder in full. The prosecutor determines the probative value of the evidence, what it can or cannot be used to prove. The prosecutor’s decisions almost never take into account a presumption of innocence or a felt need to prove guilt beyond a reasonable doubt. The job of adjudicating guilt and fixing punishment is effectively transferred from neutral courts to partisan prosecutors. The judicial process, with its ancient institutions and ideological commitments, is transformed into an administrative system with very different objectives and standards. Because this outcome of plea bargaining is endemic rather than idiosyncratic, a consequence of the system working routinely – as it is supposed to do – in the majority of cases rather than the occasional malfunction or malfeasance, the proper claim would seem to be not that any particular plea bargain subverts the defendant’s right to be presumed innocent, a claim that would present formidable problems of proof in any case, but that the system itself operates in violation of s 35(3)(h). Taken together, its structure of powers, incentives, and outcomes is inconsistent with reserving conviction and punishment solely for defendants who have been convicted beyond a reasonable doubt. The remedy would thus not be reopening the occasional bad bargain, as in Santobello, but overturning the system of plea bargaining itself.

A reply to this claim might be grounded in Uijs J’s analogy of plea bargains to civil settlements. In civil cases, where the stakes for the defendant do not involve incarceration and the plaintiff’s burden of proof is less onerous, the parties are given full freedom to determine by negotiation the precise distribution of liability and damages that the court will be asked to ratify, and their decisions in bargaining may be influenced by whatever particular interest or objective they may have in the lawsuit. It is the plaintiff’s right to sue, and the defendant’s right to choose whether or not to defend in court. Criminal proceedings, the argument might go, should be seen no differently. Inquisitorial criminal processes may have a rule of compulsory prosecution, a requirement that, to the extent possible, every offender should be prosecuted and punished for the highest crime the evidence in the case will support. Adversarial systems like South Africa’s do not. The same ancient traditions to which the claim appeals also give prosecutors substantial discretion to select charges not solely on the basis of factual accuracy but on their view of what justice requires as well. More importantly, defendants, who may have many reasons for wanting to settle a criminal case ‘out of court’, have always had the right to do so if they wish by pleading guilty. To overturn the plea bargaining system would require the withdrawal of both the prosecutor’s charging discretion and the defendant’s right to plead guilty, and go far toward transforming the adversarial system on which the Constitution’s array of powers and rights is based into an inquisitorial system inconsistent with its common law and constitutional values.

Plea bargaining’s opponents might seize on this last point, but turn it around. Even where formal agreements are reached and ratified under s 105A, they might claim, the system itself replaces adversarial procedures of accusation and proof before an impartial judge with administrative resolution by prosecutors, informed by evidence gathered by police and endorsed by compliant judges after cursory inquiries into the facts. It is an inquisitorial system in all but name, without the protections afforded by the rule of compulsory prosecution and the conduct of investigative trials by forensically-trained judges professionally committed to finding the truth in every case before them. In the twilight regime of conviction by agreement, halfway between the adversarial and the inquisitorial, plea bargains may be precisely the ‘summary proceedings which bear little resemblance to criminal trials as [the] main process to incarcerate
those suspected of crimes’ that Snyckers and le Roux feared might follow from \textit{Nel} and afford defendants even fewer protections than those offered by s 35.\footnote{Snyckers and le Roux (note 70 above), at 51-13.} If so, and an ‘ordinary court’ is held to be an adversarial court, then the system of plea bargaining might be challenged under s 35(3)(c) of the Constitution for failing to guarantee defendants ‘a public trial in an ordinary court’. Despite the reasoning in \textit{Nel}, and despite the legislative approval of plea bargaining signaled by passage of s 105A, Snyckers and le Roux insist that

The clear aim of this provision is to ensure that normal criminal trials with full fair-trial protection are the only acceptable method whereby the state may prosecute individuals for committing offences. Exceptions to this rule require FC s 36(1) justification.\footnote{Ibid at 51–94.}

If this is held to be the case, and the regime of plea bargaining is found to infringe either the right of defendants to the presumption of innocence under s 35(3)(h) of the Constitution or a fair public trial under s 35(3)(c), might these limitations themselves be justified? Once again, no more than a suggestive sketch can be offered here, but the broad outlines of such a justification in the universal problem of high caseloads and scarce judicial resources seem clear enough. Section 36(1) of the Constitution provides that

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.\footnote{Constitution of the Republic of South Africa, 1996, s 36(1).}

The limitation in question would be the general infringement of the rights of defendants under s 35(3)(h) or s 35(3)(c) by the system of plea bargaining operating within the procedures provided by ss 112, 113 and 105A of the Criminal Procedure Act 51 of 1977 as amended – all laws of general application. To support the claim that the limitations the system places on these rights are reasonable and justifiable in terms of the underlying values expressed in the section, the state might address each of the five factors in turn.

It cannot be denied that access to a fair trial with a presumption of innocence is a right of first importance to an open and democratic society based on human dignity, equality and freedom. However, a fair adversarial trial ‘with trimmings’ is an expensive proposition, one that must be paid for in largest part from the public purse. Plea bargaining is an attempt to direct extremely scarce judicial resources toward achieving convictions and some measure of punishment where appropriate in the greatest number of cases possible, without which the public’s dissatisfaction with the operation of criminal justice might be substantially aggravated. This limitation, importantly, operates with the lightest possible touch on defendants. It does not require them to surrender any rights; it merely tries to make it attractive for them to do so voluntarily. Every defendant who insists on a trial will have one, even if the sentencing discounts that the plea bargaining system offers lead few to want one. The purpose of the limitations that the system of plea bargaining places on the rights of defendants is to conserve not just the material resources of the criminal process, but at the same time to preserve the ability of the process
itself to provide a satisfactory resolution to as many cases of crime as possible, a public good of highest importance to societies everywhere. A court might well conclude that it achieves all of these ends in a manner that evolves ‘naturally’ from the adversarial structure itself, one that does less violence to it than would an imperfect implantation of inquisitorial procedures and assumptions. Finally, since both the courts and the legislature have acknowledged that adequate funding cannot be found to meet the cost of providing every defendant with a full criminal trial, plea bargaining can plausibly be said to be the least restrictive means of achieving the goal of justice in an open and democratic society based upon human dignity, equality and freedom.

As the Constitutional Court has repeatedly held, ‘the least restrictive means’ to limit a constitutional right does not, and must not, oblige the Court to substitute its judgment of what it ‘imagines’ the fairest possible arrangement to be for the legislature’s well-considered judgment regarding the ‘least restrictive means’. Woolman and Botha note that ‘[in State v] Manamela, the Court displays [an] awareness of the difficulties associated with determining the availability of less restrictive means, [and] arrives at a somewhat crisper articulation of the problem’ that the phrase suggests.80 “The problem for the Court is to give meaning and effect to the factor of less restrictive means without unduly narrowing the range of policy choices available to the Legislature in a specific area”.81 The Manamela Court goes on to observe:

It will often be possible for a court to conceive of less restrictive means, as Blackmun J has tellingly observed: ‘And for me, “least drastic means” is a slippery slope ... 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 legislation down’.82

The Manamela Court’s fear that judges may ‘annihilate the range of choices’ available to the legislature flows from its recognition that the phrase ‘less restrictive means’ often raises difficult issues relating to ‘cost, practical implementation, the prioritisation of certain social demands and the need to reconcile conflicting interests’.83 In State v Mamabolo, Kriegler J addresses these concerns in the following manner:

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.84

The same fears underlie the concerns that many have expressed about the use of plea bargains in an adversarial system constitutionally committed to fair trials. However, as I have tried to show here, depending on the circumstances that make them necessary, the range of ‘less restrictive’ institutional means to limit a defendant’s constitutional right ‘to a public trial before an ordinary court’ may be quite wide. In South Africa, both the courts and the legislature have attempted to ensure that defendants in the nation’s criminal courts are able to make reasonably informed decisions at the same time that the public sees justice as being broadly served in the dispensation of punishments. In this way, plea bargains, in all their complexity,

80 Woolman and Botha (note 75 above) at 88–89.
81 S v Manamela & Another (Director-General of Justice Intervening) 2000 (3) SA 1 (CC), 2000 (5) BCLR 491 (CC) at para 95.
83 Ibid at para 95.
84 S v Mamabolo 2001 (3) SA 409 (CC), 2001 (5) BCLR 449 (CC) at para 49.
can persuasively be said to preserve the system of justice upon which South Africa’s open and
democratic society relies. They will almost certainly continue to do so until the nation possesses
the resources necessary to offer a full trial to every criminal defendant.

Economic ways of thinking do not provide answers to questions such as these so much as
they provide useful ways for people to think, and disagree, about them. I do not presume to
have offered more than this, or to urge any particular resolution of the constitutional issues
raised by plea bargaining. Making and defending the kinds of constitutional claims sketched
here would surely require not just more imaginative lawyering than this, but a good deal of
sophisticated empirical research on the environmental and material conditions of criminal
justice in South Africa, the incidence of guilty pleas, both formal and informal, in the system,
and the sentences actually being imposed after both guilty pleas and convictions at trial. All of
these are tasks for another time. I hope here simply to have opened new avenues of study and
debate about a problem of great importance to the law in South Africa, and around the world.
Reconciliation as the Aim of a Criminal Trial: 
Ubuntu’s Implications for Sentencing

THADDEUS METZ

ABSTRACT: In this article, I seek to answer the following questions: What would a characteristically African, and specifically relational, conception of a criminal trial’s final end look like? What would the Afro-relational approach prescribe for sentencing? Would its implications for this matter forcefully rival the kinds of penalties that judges in South Africa and similar jurisdictions typically mete out? After pointing out how the southern African ethic of ubuntu is well understood as a relational ethic, I draw out of it a conception of reconciliation that I advance as a strong candidate for being the proper final end of a criminal trial. I argue that, far from requiring forgiveness, seeking reconciliation can provide strong reason to punish offenders. Specifically, a reconciliatory sentence is one that roughly has offenders reform their characters and compensate their victims in ways the offenders find burdensome, thereby disavowing the crime and tending to foster cooperation and mutual aid. I argue that this novel account of punishment is a prima facie attractive alternative to more familiar retributive and deterrence rationales, and that it entails that widespread practices such as imprisonment and mandatory minimum sentences are unjust.

KEYWORDS: African jurisprudence, communal relationship, criminal justice, imprisonment, punishment, reconciliation, sentencing, sub-saharan morality, ubuntu

AUTHOR: Distinguished Professor, University of Johannesburg.
Email: tmetz@uj.ac.za

ACKNOWLEDGEMENTS: This article has been improved as a result of oral feedback received from participants at the Constitutional Court Review IX Conference and written feedback from Lucy Allias, Stephen Kershnar, Stu Woolman and two anonymous reviewers.
I INTRODUCTION

It might seem obvious what the point of a criminal trial is: to ascertain whether there is legal guilt and to impose a penalty in response to it. However, from a philosophical perspective, it is fair to ask why judges should impose a penalty in response to a crime, not to mention which penalties are appropriate. It is not enough to respond that judges should impose certain penalties because the common law or a legislative or constitutional provision instructs them to, for that begs the question of why there should be law requiring judges to respond to crimes in those ways.

Salient accounts of what judges should ultimately be trying to achieve when responding to crime in contemporary English-speaking jurisprudence include protecting rights, giving people what they deserve and other views that are Western and individualist (in senses spelled out below). These conceptions of the proper final end of a trial have important ramifications for which penalties, if any, a judge ought to impose.

In this article, I seek to answer the following cluster of questions: What would a characteristically African, and specifically relational, conception of a criminal trial’s final end look like? What would the Afro-relational approach prescribe for sentencing? Would its implications forcefully rival the kinds of penalties that judges in South Africa and similar jurisdictions typically mete out?

In what follows I answer these questions about respects in which a judge should sentence offenders in the light of ubuntu, the Nguni term meaning humanness that is widely used in South Africa and more broadly on the continent to capture indigenous sub-Saharan ideas about morality. After pointing out how ubuntu is well understood as a relational ethic, I draw out of it a certain conception of reconciliation that I advance as a strong candidate for being the proper final end of a criminal trial and as having prima facie attractive implications for how to punish offenders. Although a reconciliatory approach to sentencing is from (South) Africa, it is not meant to be only for (South) Africa; those working in other traditions should be able to find something of prima facie interest in it.

Of course, others are known for having argued that ubuntu and related traditional African values prescribe reconciliation, social cohesion or restorative justice. However, when such an approach has been applied to contemporary, large-scale societies, it has usually been viewed as an alternative dispute resolution mechanism, that is, as something to be sought only in the cases of adolescents, or less serious offences, or transitional societies, or customary law matters. In contrast, I am interested in how ubuntu might ground a mainstream approach to the way judges in ‘modern’ societies punitively respond to violations of criminal law, which has yet to be considered.

---


2 Dikoko v Mokhatla [2006] ZACC 10 at para 115 (Sachs J points out disapprovingly). (Dikoko)


4 D Tutu No Future without Forgiveness (1999).

In addition, others have contended that reconciliation essentially requires forgiveness of offenders or otherwise demands a non-punitive response to them if at all possible. However, I argue that an ubuntu-based conception of reconciliation between disputants as the point of a criminal trial grounds a novel account of when and how judges should punish offenders, one that is a real competitor to dominant approaches to sentencing. In a nutshell, instead of seeking to give offenders what they deserve or to instil fear in prospective offenders so as to prevent rights violations, reconciliation would have offenders reform their characters and compensate their victims in ways the offenders find burdensome, thereby disavowing the crime while tending to foster cooperation and mutual aid.

In order to illustrate the differences between these approaches to punishment, I address the recent South African Constitutional Court case of Ndlovu v The State. The case was about precisely which mandatory minimum prison sentence to impose on Brendan Solly Ndlovu, who was morally guilty of having both committed rape and inflicted serious bodily harm, but had been charged with only the former. Ndlovu naturally preferred a lighter sentence than what the state sought. I, however, provide reason to doubt both sides of the dispute, maintaining that no mandatory minimum prison sentence would have been ideal, even given Ndlovu’s serious crimes. Although mandatory minimums could make sense by appeal to retribution or deterrence, I argue that they are largely out of place in a reconciliatory scheme, by which judges would routinely need to attend to the specifics of the offender, his victim and the broader social context. Furthermore, I contend that while punishing with imprisonment could well be prescribed by considerations of retribution or deterrence, doing so is normally proscribed by a reconciliatory perspective.

In the next section of this article, I begin by reminding readers of the most influential conceptions of the point of a criminal trial, noting how it entails certain conceptions of just punishment, and indicating why these conceptions count as Western and individualist (section II). Then, I sketch a moral-philosophical interpretation of the African ethic of ubuntu and draw from it a certain conception of reconciliation (section III). In the next section, I suppose that the point of a criminal trial is to bring about such reconciliation and spell out what this would mean for how to punish (section IV). Along the way, I contrast reconciliatory sentencing with the more Western/individualist theories of punishment as well as indicate how reconciliatory sentencing prescribes changes to some current practices that these theories support using Ndlovu as a foil. I conclude by noting some further research that it would make sense to undertake. This includes how to obtain evidence of guilt in a criminal trial and what to criminalise in the first place, supposing this article has advanced a plausible, Afro-centric and reconciliatory alternative to dominant models of criminal justice (section V).

II DOMINANT PHILOSOPHIES OF CRIMINAL JUSTICE

In this section I briefly recount the philosophies of criminal justice that have been particularly prominent in English-speaking jurisprudence and have influenced practice in South Africa and similar jurisdictions in both Africa and the West the most. I also explain why I label

---

them ‘Western’ and ‘individualist’, in contrast to the more ‘African’ and ‘relational’ approach articulated in the following section.

By ‘criminal justice’ I mean the contested issue of how the political community should respond morally to legal transgressions, where one potential response is the infliction of punishment. State punishment, in turn, is essentially a matter of trying to burden or deprive someone in response to a legal transgression that has been committed (or represented as such). So defined, a criminal justice system that includes state punishment differs from civil justice, defensive force and quarantine, which do not necessarily include aiming to burden someone because of a legal transgression.

Retributivism, or the ‘pay back’ theory, is one influential account of why there should be a criminal trial and of how state punishment should be inflicted to achieve its final end. I use the term ‘retributivism’ broadly to signify the view that the aim of a trial should be to determine whether someone is guilty of having broken a just law and to impose a penalty that is proportionate to the crime he committed in the past. In the South African context, this approach is often alluded to with the suggestion that sentencing ought to fit ‘the nature of the crime’ or ‘the seriousness of the offence’ and that there is a presumption against ‘disturbing disparities’ or a ‘striking difference’ between two sentences for the same crime. Someone guilty of rape, such as Ndlovu, should invariably receive a harsh penalty, much harsher than someone guilty of stealing a candy bar, and just because of what they have respectively done.

Within the genus of retributivism, philosophers of law have articulated various species. For instance, there is desert theory, the view that the aim of a criminal trial should be to give offenders what they deserve for having culpably done wrong. Just as one can positively deserve a job, a raise or a reward of a sort that is proportionate to having done well, so one can negatively deserve a penalty that is fitting in respect of having acted poorly.

Another instance of retributivism is fairness theory, the view that the point of a criminal trial is to ascertain which, if any, crime occurred and to impose a penalty that will remove the unfair advantage the criminal thereby took of other law-abiding residents. In order for there to be rule of law, everyone must restrict their liberty by obeying it, such that when someone breaks the law, she is getting the benefit of the rule of law without undergoing the burden needed to produce it. Punishment is thought to remove the unfairness by imposing a burden similar in degree to that which the offender failed to carry.

Retributivism is a ‘backward-looking’ view, directing a judge to consider the past in order to ascertain what to do in the present. When imposing a sentence, a judge has to consider whether a defendant is guilty of a crime and, if so, how grave a crime it was, and from that determine which penalty is proportionate to the severity of the crime. The only consideration about the future a retributivist might routinely consider is whether by imposing a penalty on an offender the court would foreseeably bring undeserved harm to

---

10 J Murphy Retribution, Justice, and Therapy (1979); M Davis To Make the Punishment Fit the Crime (1992); and R Dagger ‘Playing Fair with Punishment’ (1993) 103 Ethics 473.
some other innocent parties such as his children. Otherwise, a retributivist does not believe that the point of a sentence is to prevent either the offender or others in society from committing crime down the road.

In contrast, the other familiar class of accounts of criminal justice are resolutely ‘forward-looking’, requiring the prevention of crime or similarly desirable outcomes in order for state punishment to be justified. Utilitarianism is one instance of this general category. According to this approach, the time, labour and other costs of a criminal trial are justified insofar as it would do some long-term good for society. In particular, a trial ought to put a judge in a position to ascertain whether imposing a penalty would reduce other, greater harms (bads) such as crimes, or produce compensating benefits (goods), for people more than any non-punitive response.\(^{12}\)

However, utilitarianism is not the only forward-looking view, with many of those who believe in fundamental rights also maintaining that punishment must have some desirable effect in order to be just. Consider, for instance, the idea that by having committed a crime, a criminal forfeits his right not to be punished, and that he may justly be punished if and only if doing so would protect society from similar crimes in the future.\(^{13}\) By this approach, punishment is justified in more or less the way force used in self- and other defence is, with the point of a trial being to ascertain whether someone has aggressed against others and thereby forfeited his rights and, if so, whether a penalty of some kind (no greater than the rights forfeited) would prevent future aggression either by him or others.

For both utilitarian and defensive force approaches, incapacitation and deterrence are central (even if not exhaustive) mechanisms by which punishment is thought to be desirable for controlling crime. Adherents to these influential forward-looking views welcome the prospect of potential (re-)offenders either being rendered unable to break the law or being afraid of what would happen if they exercised their ability to do so. Putting the likes of a rapist in prison for a long time could well serve both functions.

Although these are not the only accounts of why a punishment system should be established and maintained, they have been the most influential ones over the past two hundred or so years in English-speaking philosophy. They have grown out of the moral-philosophical soil that has been prominent in the West, including the duty-based principle of respect for persons articulated by Immanuel Kant; the natural rights ethic often ascribed to John Locke; and the utilitarian morality of Jeremy Bentham. It is in this sense that I call these theories of criminal justice ‘Western’; they have been salient in the philosophical thought, and also juristic practice, of North America, the United Kingdom and Europe. These accounts have been prominent in

---


many Western cultures for a long time in ways they have not in others, such as the East Asian, Indian, Middle Eastern and African.\footnote{In using the geographical label ‘Western’ in this way, I am not implying that considerations of desert, fairness, utility and defensive force have decidedly influenced thought about criminal justice either everywhere in that location or only in it. Of course there have been some Western philosophers who have rejected all these accounts, and there have been some non-Western ones who have accepted some of them. The label is meant to designate what has been particularly recurrent (not universal) in much of a locale for a long while, relative to many (not all) other spaces and times. Hence, it is quite possible that a person who lives in Africa in the twenty-first century might not hold values that are below described as ‘African’. That is, they may not subscribe to an ethos that possessed moral salience over much of the continent for several hundred years. For some defence of this way of using geographical labels, see T Metz ‘How the West Was One: The Western as Individualist, the African as Communitarian’ (2015) 47 Educational Philosophy and Theory 1175.}

Despite the important distinction between the backward- and forward-looking rationales for a criminal trial that have been prominent in the West, both share a common feature, namely, individualism. The accounts sketched above are all grounded on the idea that what gives people a moral status are features internal to them. For the Kantian it is our capacity to reason; for the Lockean it is that we own ourselves and what we work on; and for the Benthamite it is that we can feel pleasure or otherwise be satisfied. Crimes are typically viewed as degrading to an individual’s ability to make a choice for herself, or as violating her rights to control her mind, body and what she has put herself into, or as causing unnecessary pain, with penalties being justified as ways of respecting an individual’s choice, protecting individual rights to life, liberty and property, or preventing pain from coming to individuals (even if treated as a sum).

In the next section, I spell out an ethic that, in contrast, is characteristic of African cultures and is relational. From this perspective, what gives people dignity is roughly that they could interact cohesively, which grounds an approach to criminal justice according to which its aim should be to fix broken relationships, often by employing punishment.

III UBUNTU AS A RELATIONAL MORAL PHILOSOPHY\footnote{This section relies on ideas previously articulated in T Metz ‘South Africa’s Truth and Reconciliation Commission in the Light of Ubuntu’ in M Swart & K van Marle (eds) The Limits of Transition: The South African Truth and Reconciliation Commission 20 Years on (2017) 221.}

According to one large swathe of southern, and more generally sub-Saharan, African thought about morality, one’s basic goal in life should be to become a real person or develop human excellence, that is, to exhibit \textit{ubuntu}, with the central (if not sole) way to do so being roughly to commune or harmonise with others. After spelling out one plausible interpretation of what communion or harmony involves, I suggest that reconciliation is well understood as partial communion, a stepping-stone towards a fuller sort that would be ideal.

A An Afro-communal ethic

As Desmond Tutu, the influential South African Nobel Peace Prize winner and Chairperson of the Truth and Reconciliation Commission, has remarked of characteristically African approaches to morality –

‘We say, ‘a person is a person through other people’. It is not ‘I think therefore I am’. It says rather: ‘I am human because I belong,’ I participate, I share … Harmony, friendliness, community are great goods. Social harmony is for us the \textit{summum bonum} – the greatest good.’\footnote{Tutu (note 4 above) 35.}
As for what a harmonious or communal relationship involves in some more detail, consider the following statements from a variety of additional South African thinkers.

Former South African Constitutional Court Justice Yvonne Mokgoro remarks in an essay on ubuntu and the law, ‘Harmony is achieved through close and sympathetic social relations within the group – thus the notion umuntu ngumuntu ngabantu/motho ke motho ka batho ba bangwe (a person is a person through other persons—ed.’. Gessler Muxe Nkondo, who has had positions of leadership on South Africa’s National Heritage Council, says that ‘ubuntu advocates … express commitment to the good of the community in which their identities were formed, and a need to experience their lives as bound up in that of their community’. Nhlanhla Mkhize, an academic psychologist and Dean at the University of KwaZulu-Natal, remarks that ‘personhood is defined in relation to the community … A sense of community exists if people are mutually responsive to one another’s needs … [O]ne attains the complements associated with full or mature selfhood through participation in a community of similarly constituted selves … To be is to belong and to participate’. For a final example, two South African theologians, Mluleki Mnyaka and Mokgethi Motlhabi, understand ubuntu as follows:

Indivduals consider themselves integral parts of the whole community. A person is socialised to think of himself, or herself, as inextricably bound to others … Ubuntu ethics can be termed anti-egoistic as it (sic) discourages people from seeking their own good without regard for, or to the detriment of, others and the community.

This characterization of ubuntu and other construals of communing or living harmoniously with others embrace two recurrent themes. On the one hand, part of so relating is what I call ‘identifying with others’ or ‘sharing a way of life’, a matter of participating, being close, experiencing life as bound up with others, belonging and considering oneself a part of the whole. On the other hand, one finds references to sharing, being sympathetic, being committed to others, responding to others’ needs and acting for others’ good, which I call ‘exhibiting solidarity’ or ‘caring’.

More carefully, it is revealing to understand identifying with another (or being close, belonging, etc) to be the combination of exhibiting certain psychological attitudes of we-ness and cooperative behaviour. The psychological attitudes include a tendency to think of oneself as a member of a group with the other and to refer to oneself as a ‘we’ (rather than an ‘I’), a disposition to feel pride or shame in what the other or one’s group does, and, at a higher level of intensity, an emotional appreciation of the other’s nature and value. The cooperative behaviours include being transparent about the terms of interaction, allowing others to make voluntary choices, acting on the basis of trust, engaging in joint projects, and, at the extreme end, choosing for the reason that ‘this is who we are’.

The other part of communing or harmonising, namely, exhibiting solidarity with another (or acting for others’ good, etc), is also usefully construed as the combination of exhibiting

---

certain psychological attitudes and engaging in certain kinds of behaviour. Here, the attitudes are ones positively oriented towards the other’s quality of life, and they include a belief that the other merits aid for her own sake, an empathetic awareness of the other’s condition, and a sympathetic emotional reaction to this awareness. The actions are not merely those likely to be beneficial, that is, to make the other better off in welfarist terms, but also, in the ideal case, are ones done for that reason and for the sake of making the other a better person or for the sake of communal relationship itself.

Communal relationship

Identity
(share way of life)

Sense of togetherness

Solidarity
(care for quality of life)

Cooperative participation

Aid

Sympathetic altruism

Figure 1: Schematic representation of communion

In sum, what I take to be the attractive moral core of ubuntu is the proposition that one ought to develop human excellence, which one can do by sharing a way of life with others and caring for them, or, more carefully, by treating people as having a dignity in virtue of their capacity to be party to these communal or harmonious relationships. This analysis of ubuntu is a moral-philosophical reconstruction, a secular one that does not rely on highly contested metaphysical claims about the existence of imperceptible agents such as ancestors and forces such as witchcraft, and one that should be prima facie appealing to readers from a wide array of cultures.

By this approach to ethics, any given agent should, roughly, seek to create, maintain and enrich relationships of sharing and caring. Conversely, one would act wrongly if one either failed to do so (at least with other innocent parties), or, worse, prized the opposite, anti-social relationships. On the face of it, actions that intuitively count as crimes are ones in which people act out of an ‘us versus them’ attitude, subordinate others, harm them and do so out of indifference to their good or even cruelty. In terms of ubuntu, it is relating discordantly in these ways that calls for a reconciliatory response.

One can see how an interest in reconciliation consequent to crime would follow from an ethic of treating people with respect insofar as they are capable of identity and solidarity. If what is of utmost importance is relating communally or harmoniously, then one who acts in a discordant manner should be treated in ways that are likely to counteract his discordance and to foster harmony between him and others. The focus on rebuilding relationships differs from an individualist focus on, say, giving people what they deserve or preventing pain from coming to them, with further contrasts drawn below.
B  Reconciliation as a function of communion

In this section, I appeal to elements of the African ideal of communion to articulate the essentials of a desirable kind of reconciliation. It is only in the following section that I draw from it a certain punitive approach to responding to violations of criminal law.

In general, reconciliation is something that occurs after a period of conflict, ranging from a fight between spouses to systematic human rights violations between groups. In addition, while reconciliation is probably good for its own sake to some degree, few think of it as an ideal, with reconciliation instead usually being understood as something that lays a path towards an ideal. At a social level, reconciliation is not necessarily full-blown distributive justice, and at the interpersonal level it is not essentially real love.

There are a variety of ways of conceiving of a desirable, but less than ideal, form of social interaction consequent to conflict. The one I propose is grounded on the ubuntu ethic sketched above. Suppose, with ubuntu, that a state’s moral duty (or that of its officials, if one is sceptical that a state as such is a collective agent that has moral duties) is to treat residents as special in virtue of their capacity for communal/harmonious relationships. One clear way of doing so would be for a state to foster such relationships between people in its territory, that is, to actualise the special capacity. Now, if one proper final end of the state were to engender communal relationships, and if reconciliation were a stepping stone towards such a condition, then it would be sensible to think of reconciliation as constituted by some of the elements of communion.

Specifically, a promising conception of reconciliation is based mainly on what I labelled the behavioural facets of a characteristically African conception of communion/harmony, and not as much on the attitudinal ones. As a first approximation, consider the view that to reconcile is for two parties to engage in cooperative behaviour oriented towards mutual aid. Such need not involve mental states such as thinking of oneself as a ‘we’, taking pride in others’ accomplishments, exhibiting sympathy towards others or acting for their sake. Of course, people’s hearts and minds would need to change to some degree in order to move from serious discord, roughly substantial subordination and harm, to a way of relating with the core, behavioural components of identity and solidarity as above. However, they would need to do so to a much lesser degree than they would in order, say, to be motivated by altruism or compassion, or to enjoy a sense of togetherness. Supposing, then, that communion is central to ethics and that reconciliation contributes towards full-blown communion, it is plausible to

---


23 Metz ‘A Theory of National Reconciliation’ (note 22 above) 120–121.

24 Were we to require inclusion of these psychological responses, we would be expecting ‘too much’ from the concept of reconciliation, and demanding that it conform too closely to a specific and rather demanding social ideal. After a period of serious conflict between people, one cannot expect their beliefs and emotions to change quickly, whereas their behaviour might. Consider that although after World War II many West Germans reportedly continued to favour Hitler’s policies, they nonetheless conformed to a constitutional order that sought to repair some of the damage done to the Jewish and other oppressed populations. Whatever (desirable) changes to the collective mindset that (ostensibly) occurred, they came much later. What goes for two populations might plausibly be said to hold for two people.
think of reconciliation as consisting mainly of the behavioural facets of it, and not requiring all the emotional and motivational ones.

Note that, unlike some conceptions of reconciliation offered by South African thinkers, particularly those in the Christian tradition, forgiveness, understood as including the dissipation of negative emotions, is not essential to the conception I propose. Neither is empathy, nor is ‘a spiritual sense of belonging and community that draws people towards a fullness of humanity through others’. Forgiveness, empathy and a spiritual sense of belonging would be elements of a complete communion (or perhaps the very best form of reconciliation), but are too demanding for reconciliation as such, which is less than ideal.

One way to honour the final value of people’s capacity for communion is to realise it as an end, that is, to actualise communal relationships; however, that is not the only way to express certain positive attitudes towards this special capacity. Specifically, an attractive notion of reconciliation, as a way of respecting people’s capacity to commune, is one that also includes the disavowal of disrespectful treatments of this superlative value that were undertaken in the past. That would involve those associated with victims, such as family members and wrongdoers, as well as the political community in certain kinds of cases, expressing disapproval of the wrongdoing (ie, roughly the pricing of discord) that took place.

Sometimes people who have been party to a conflict that includes wrongdoing are able to come together and repair the relationship without thinking in those terms, perhaps electing to forget without any moral reflection on what transpired. However, my suggestion is that usually such so-called ‘reconciliation’ is not particularly desirable. To honour the value of communion means acknowledging when it has been seriously undermined in impermissible ways (when it comes to victims); and to treat people as special by virtue of their capacity to commune also means responding to offenders in the light of the way they have misused this capacity. In the best case scenario, this would involve offenders hearing victims out, apologising to them and expressing remorse by striving both to make up for wrongful harm done and to avoid repeating the wrong. However, at least where offenders are unwilling to do these things, a political community that has taken responsibility for upholding residents’ dignity should hear victims out, acknowledge aspects in which they were mistreated, express disapproval of how they were mistreated and do so by effecting reparations for wrongful harm they suffered and making it clear that it will protect them from further mistreatment.

Putting things together, here is an ubuntu-based account of an attractive conception of reconciliation: a condition consequent to interpersonal conflict in which those directly affected by it interact on a largely voluntary, transparent and trustworthy basis for the sake of compossible ends largely oriented towards doing what is expected to be good for one another and in which those associated with victims disavow wrongdoing that was part of the conflict. Notice the two basic parts here: the realisation of (behavioural) harmony and the disavowal of prior discord. Mere disavowal of a wrong would not be sufficient for something to count as reconciliation at all, while mere (behavioural) harmony consequent to a wrong would be sufficient for reconciliation, but not a particularly welcome form of it. If, by ubuntu, we must

25 Several authors cited near the beginning of this article (note 6 above) offer good examples of this mindset.
treat people as special by virtue of their capacity for harmonious relationships, then both facets are plausibly essential for a desirable kind of reconciliation.

IV RECONCILIATORY SENTENCING

In this section, I spell out what it would mean for a criminal trial to seek out reconciliation as an end, focusing in particular on the implications of reconciliation for laying down penalties. I contrast reconciliatory prescriptions for sentencing both with Western-individualist theories of punishment and with some predominant ways that people are punished in South Africa and similar jurisdictions, including imprisonment and mandatory minimum sentences. Although I do not aim to convince the reader that a reconciliatory approach to sentencing is justified, I do try to show that it should not be dismissed as a rival to more familiar models and practices.

A Reconciliation as fact-finding

If reconciliation between at least offenders and victims were a central point of a criminal trial, then an important task for a judge would be to obtain the truth about what had transpired. For one, the notion of reconciliation from the previous section requires disavowing wrongful discord, which, in turn, means that there is a clear fact-finding role for a court: it must discover who acted wrongfully and in what respects. A court must sort between the innocent and the guilty, and ascertain how much guilt there is and for what, so as to be in a position to express disapproval of wrongdoing.

Furthermore, even the advancement of behavioural communion, in the form of cooperation, prescribes truth about the past. Two parties genuinely share a way of life when they are clear about how they have interacted in serious ways that potentially include wrongdoing, or at least when they are aware of what the other thinks about that. And those associated with these two parties, such as family members and co-workers, also need the truth about the past in order to share a way of life with them, as opposed to being isolated by virtue of ignorance.

It is an empirical question of whether an adversarial system or an inquisitorial one (or a mix, or something else) would do a satisfactory job of revealing the truth about past wrongdoing. Legal scholars who draw on traditional African culture have invariably eschewed adversarialism, since its competitive or combative nature appears incompatible with duties to do what is best for society or to foster reconciliation amongst disputants. However, if reconciliation includes moving forward together in the light of an accurate awareness of what transpired in the past, and if it happened to be the case that an adversarial system is necessary to facilitate that adequately or did so to a much greater degree, then there would be a strong, under-appreciated case for adversarialism on grounds of reconciliation. Any plausible African ethic will make space for competitive fora such as sports and markets (which does not necessarily mean capitalism) roughly because they can facilitate a greater harmony on balance for society, and it could be that a competitive courtroom is analogously justified.

28 In the conclusion I raise the issue of whether reconciliation should be sought amongst more than just victims and offenders.
30 Idowu (note 1 above) 13, 15.
31 In any event, how to obtain evidence of guilt is an issue to address in detail elsewhere. The focus of this article is how to punish persons once one has obtained evidence of their guilt.
B Reconciliation as punishment

Upon having ascertained that there was a crime and who committed it, a judge must decide how the state should respond to the criminal. As noted in the previous section, the sort of reconciliation that *ubuntu* (as understood here) prescribes does not require forgiveness. Letting go of resentment and related negative attitudes towards an offender could sometimes be instrumentally useful for enabling reconciliation, but reconciliation, as the combination of the realisation of (behavioural) harmony and the disavowal of prior discord, does not essentially consist of that or otherwise require it.

In addition, I now seek to rebut the widespread presumption that reconciliation or restorative justice is incompatible with punishment. Justice Sachs’ discussion in *Dikoko* suggests such a view. According to him, *ubuntu* prescribes restorative justice, where ‘the key elements of restorative justice have been identified as encounter, reparation, reintegration and participation’ and where reparation ‘focuses on repairing the harm that has been done rather than on doling out punishment’

Although it is doubtful that Sachs would eschew punishment altogether in a criminal justice system given what he says about the need for deterrence, the thrust of his remarks suggest that restoring relationships, including by compensating victims for harm they have undergone, is an alternative to punishment. In what follows I argue that an *ubuntu*-based reconciliation often prescribes punishment, indeed as a way to compensate victims.

One tempting strategy by which to show that reconciliation can prescribe punishment is to note that there are situations in which reconciliation between victims and offenders would be possible only after victims were satisfied that offenders had been punished in retributive fashion. This approach has been suggested by the social scientist Brandon Hamber, informed by his engagements with victims of apartheid-era political crimes. He and various co-investigators found that ‘there remains a strong feeling amongst victims/survivors that justice should be done and that this is necessary if we are to create a new society’. Similarly, Hamber and others remark, ‘The door to reconciliation and forgiveness will be opened all that wider if the desire for revenge is legitimised and understood, if it is respected and contained, and if it is given both public and private space for its expression’. Basically the idea is that only upon seeing offenders receive their just deserts would victims be likely to accept their reintegration into society.

---

32 *Dikoko* (note 2 above) at para 114.
33 *Dikoko* (note 2 above) at para 114. Elsewhere in this judgment, Justice Sachs remarks: ‘The principal goal should be repair rather than punishment. To achieve this objective requires making greater allowance in defamation proceedings for acknowledging the constitutional values of *ubuntu* – botho’. Ibid at para 112.
34 *Dikoko* (note 2 above) at para 120.
However, this is not my preferred approach, since the connection between what is called ‘reconciliation’ and punishment is not strong enough. By the above rationale, punishment would be unjustified if it were unnecessary for victims and offenders to ‘reconcile’ afterwards. However, many readers will share the intuition that punishment for many kinds of crimes would be justified even if victims were to forgive offenders and absolve them of deserved punishment. Similar offences should receive comparable penalties, in terms of the degree of burden involved.

In contrast, my main strategy is to contend that the preferred understanding of ‘reconciliation’ routinely includes punishment as partially constitutive of it. Instead of deeming punishment to cause reconciliation on occasion, my suggestion is that punishment (nearly) always helps to constitute reconciliation. Roughly speaking, the ubuntu-based account of reconciliation typically carries a certain kind of punitive justice within it that differs from Western retributivism.

There are two reasons for thinking so, grounded on each of the two major facets of reconciliation, viz, promotion of behavioural communion in the future and disavowal of its having been flouted in the past. First, consider reconciliation insofar as it includes behavioural communion in the form of cooperation and aid. That can prescribe compensation; making reparations to a victim would be one way for an offender to cooperate with and aid her, and often it would be a burden to do so. However, if an offender were wealthy, then making reparations would not be burdensome, making this rationale unable to ground punishment to the degree that is intuitively warranted. Similarly, if a victim were poor, an offender might be able to pay minimal compensation that would aid the victim in a way that would be welcome relative to her means, but fail to track the degree of the offence.

Relatedly, behavioural communion in the future could be advanced by preventing recidivism on the part of the offender. An offender has a duty to reform himself to avoid committing crime again, and the state has an obligation to take steps to prevent a wrongdoer from doing wrong again. That is especially true if the wrongdoer is not doing so himself or is unable to do so on his own. The Court made this connection between ubuntu and rehabilitation in S v Makwanyane:

[T]he reformative theory … considers punishment to be a means to an end, and not an end in itself – that end being the reformation of the criminal as a person, so that the person may, at a certain stage, become a normal law-abiding and useful member of the community once again …

This, in my view, accords fully with the concept of ubuntu.37 However, the same concern about this reasoning arises, namely, that reform and consequent cooperation and mutual aid could conceivably occur without hard treatment of the offender being involved; consider a spontaneous ‘come to Jesus’ moment on his part.

Therefore, essential to grounding a reconciliation-based justification of punishment is its other facet, the disavowal of wrongful discord. It is this under-appreciated, squarely expressive dimension of reconciliation that, I argue, reliably brings accountability in the form of deprivation in its wake. ‘Actions speak louder than words.’ ‘Put your money where your mouth is.’ ‘Talk is cheap.’ In addition to these maxims, a fortune cookie once told me, ‘A person of words and not of deeds is like a garden full of weeds.’ I maintain that, in cases of serious crime, reconciliation prescribes burdensome compensation and burdensome rehabilitation as ways of expressing disapproval on the part of the political community, and also, in the best case,

remorse on the part of the offender.\textsuperscript{38} For an offender merely to apologise or for a court merely to wag a disapproving finger at him would be inadequate forms of disavowal; in a word, there must also be some hardship for the disavowal of a serious crime to be meaningfully expressed, whether by the offender or the court.\textsuperscript{39}

The degree of hardship imposed should \textit{track} the degree of wrongdoing, in the sense that the worse the wrong, the greater the hardship, although retributive \textit{proportionality} is not required. For example, to express disavowal of torture adequately, a court must impose a weighty burden indeed on the torturer, but it need not sentence him to be tortured or to a fate strictly proportionate to the wrongful harm of that crime. So long as all torturers were to receive comparably stiff sentences, even if they were not a matter of torture or its exact equivalent, there would be consistency in sentencing.\textsuperscript{40}

Although compensation merely for the sake of moving forward together need not involve hard treatment of an offender, compensation in order to disavow a crime plausibly must. If an offender were truly sorry and wanted to demonstrate his guilt, he would be willing to place hardship on himself as a way to display those emotions, where the greater his wrongdoing and the stronger his apt emotional reactions to it, the heavier the hardship. Hence, if the offender were rich, he would do more than just cut a cheque to the victim. And if a court were truly disapproving of a crime, it would compel the offender to make restitution in a way that involved real labour or some other burden. Where making financial compensation would mean a change in lifestyle for an offender, it could well be a sentence that adequately disavows the offence.

There are, however, ways that compensation could place a weighty burden on an offender that are not financial. Perhaps someone who cheats on his taxes should be made to perform some dull tasks for the state revenue service. Maybe a person who has robbed a household should wear a uniform and serve as a neighbourhood-watch guard for a time. Possibly someone who has unjustifiably taken the life of a breadwinner should farm with his hands, providing sustenance to the victim’s family.

Of course, victims might not want to be in contact with offenders.\textsuperscript{41} \textit{Ubuntu} likely places some form of moral obligation upon victims to try to reconcile with offenders -- say, by accepting their offers of restitution. It does not follow that a court should force victims to do so. In such a case, victims might indicate a preferred way in which offenders should direct their efforts. After consultation with a woman who had been physically abused, for instance, a

\textsuperscript{38} In the following, I assume that the degree to which an offender should express remorse and the degree to which a political community should express disapproval align, although I recognise that this assumption may be questioned and might deserve an extended defence elsewhere.

\textsuperscript{39} Retributive expressivists have adopted this position for some time. See J Hampton ‘The Retributive Idea’ in J Hampton & J Murphy \textit{Forgiveness and Mercy} (1988) 111; and T Metz ‘Realism and the Censure Theory of Punishment’ in P Smith & P Comanducci (eds) \textit{Legal Philosophy: General Aspects} (2002) 117 (Articulates, but no longer reflects, a position I once held).

\textsuperscript{40} There are, therefore, resolutely ‘objective’ dimensions to reconciliatory sentencing: (1) facts about how bad a crime was; (2) how severe a penalty is; and (3) which penalties would track a given crime. Ultimately, a proponent of reconciliatory sentencing must provide accounts of them, but some headway can be made for now without them.

\textsuperscript{41} Although there are fascinating occasions where offenders and victims have been able to reconcile by labouring together. For an example in Canada, see ‘Convicts, Victims Work to Heal Old Wounds on B.C. Farm’ available at https://www.cbc.ca/news/canada/convicts-victims-work-to-heal-old-wounds-on-b-c-farm-13819003 (2016).
court might order her offender to undertake labour for a battered women’s shelter by delivering
needed items or helping to repair the building.

Beyond disavowing wrongful discord by ordering compensatory labour from offenders,
a court would also do so by ordering labour from offenders likely to foster moral reform.42
If offenders are genuinely remorseful, then they of their own accord would not merely take
steps, but also climb stairs, to show that they would not perform the relevant acts again. In
addition, courts would express disapproval of the wrongful behaviour by making them do so.
Such penalties would often mean mandatory therapy to get to the root of what caused the
mistreatment of others, something that would be time-consuming and psychologically difficult.
Consider as well penalties meant to instil empathy and an awareness of the consequences of
actions, such as a judge sentencing drunk drivers to work in a morgue.43 Finally, there are the
points that the hardship of punishment can often itself be a way for offenders to appreciate
how they have mistreated their victims, as well as that the guilt consequent to moral reform
would also be a foreseeable burden that offenders should undergo.

Even if an offender had a spontaneous appreciation of what he had done wrong and were
unlikely to commit similar actions again because of that, penalties would usually be apt as ways
of aiming to cultivate his moral personality while disavowing what he had done. Imagine that
you were the offender and had had a change of heart right after committing a crime. You would
want to go out of your way to show that to the victim, her family and others who reasonably
feel threatened by what you did – they could not just take your word for it. And so you would
willingly submit to burdens to express remorse, including forms of rehabilitation that would
provide all the more grounds to think that you will avoid reoffence in the future. If you were
not willing to do that, it would be right for a court to make you anyway, so as to stand up for
the victim and to censure your behaviour, all with the aim of improving relationships in the
future.

This necessity for hardship is one large difference between reconciliatory sentencing and
more familiar reformative theories of punishment.44 The latter prescribe doing something
good for the offender, helping him to become a better person, where the hardship of
punishment is conceived as the key learning tool, so that if learning were to occur without
the hardship, the logic of these theories entails that the hardship would be unjustified. In
contrast, by reconciliatory sentencing, the hardship of punishment, which is welcomed as a
potential learning tool, is inflicted additionally in order to express remorse for and disapproval
of discord (and thereby express respect for our relational nature). By this account, expressive
considerations mean that a court would usually be right to penalise offenders by ordering

42 For a useful overview of contemporary psychological and criminological research on rehabilitation, see
com/view/document/obo-9780195396607/obo-9780195396607-0046.xml
recent work in therapeutic jurisprudence, such as D Wexler ‘Therapeutic Jurisprudence’ (2004) 20 Touro Law
Review 353; and E Erez, M Kilchling & J Wemmers (eds) Therapeutic Jurisprudence and Victim Participation
therapeutic interventions, and offenders should submit to such an order, even if doing so were not in fact necessary to prevent recidivism.\footnote{This account conflicts with the principle that violence, punishment and related kinds of significant discord are justified if, and only if, they are both necessary and expected to counteract greater discord on the part of the one responsible for it. I have advanced this purely defensive account of force in previous work. T Metz ‘Human Dignity, Capital Punishment, and an African Moral Theory’ (2010) 9 Journal of Human Rights 81. In contrast, reconciliatory sentencing prescribes penalizing offenders in burdensome compensatory and reformative ways to disavow injustice, even if the burdens are unnecessary for compensation or reform. For a powerful reason to think that not all permissible uses of force are defensive, consider the case of ‘Morty’ in S Kershnar For Torture: A Rights-Based Defense (2011) 53.}

The expressive dimension of reconciliatory sentencing also enables it to make sense of why punishment would be justified in situations where improved relationships are clearly not forthcoming. For example, consider the case in which the victim has been killed and she had had no family or friends; then, no reconciliation with her or even her intimates would be possible, so that it might seem as though reconciliatory punishment would have no point. However, part of an attractive reconciliation, I have contended, is disavowing the unjust discord that took place, where offenders express remorse for what they have done and the political community expresses disapproval of it. Even if a criminal trial sometimes cannot serve the function of advancing relationships of participative cooperation and mutual aid, it could always disavow respects in which people had been wronged by the opposite, discordant ways of relating, again accounting for the intuition that consistency in sentencing must be upheld.

C Some contrasts with Western theories

In order to illustrate and motivate reconciliatory sentencing, I contrast it with the Western, individualist accounts discussed above. Although the following points are not ‘knock-down’ arguments against the latter, they provide reason to take the former seriously.

First off, it is of course a strike against a theory of punishment if it cannot explain why it is only the guilty who should be punished. Utilitarianism notoriously has difficulty restricting state punishment to those who have broken just laws, as there can be situations in which punishing people known to be innocent would (be expected to) have the best results for society. In contrast, a reconciliatory approach forbids punishment of innocent parties, since they have not done anything to undermine communal relationship. Those who have not been discordant warrant neither burdensome compensation, for there are no victims, nor burdensome rehabilitation, for no wrong has been done.

Reconciliatory sentencing avoids another famous problem for utilitarianism, namely, the imposition of disproportionately harsh penalties. In principle, severe sentences placed on a few for having committed intuitively trivial crimes could be justified if many would benefit in the long run from doing so. For example, if people risked receiving 25 years in jail for actions such as speeding, failing to indicate when changing lanes, and rolling slowly through stop signs, it could be that traffic deaths would be reduced, making the benefits to society worth the costs of occasionally ‘making an example’ out of a few offenders. However, most punishment theorists believe such penalties would be wrong, regardless of how much good they would do. Reconciliatory sentencing can account for that intuition, insofar as the degree to which the court expresses disapproval, via the imposition of burdens, should be no be greater than the...
wrongful nature of the crime (including the extent to which the criminal was responsible for it), lest the court treat the crime as more wrongful than it was.

A third advantage of reconciliatory sentencing relative to not just utilitarianism, but also other prominent forward-looking theories, is that it abjures general deterrence as a mechanism by which to control crime. Although some African theorists have appealed to general deterrence as a legitimate way to protect communal relationships, I maintain that respect for people’s capacity to commune probably forbids such an approach. If a thief wrongfully enters my house and the only way to get him to leave and without taking my things is to use a certain degree of force, I may do so. However, it would intuitively be wrongful (not merely illegal in all jurisdictions I am familiar with in North America, Europe and South Africa) to haul him out into the street and give him an additional beating intended to scare off other, potential thieves. A plausible rationale for why it would be wrong to inflict harm, such as punishment, on the guilty for the sake of general deterrence is that one is not liable for the actual or potential misdeeds of others. There is no disrespect in using substantial force, such as punishment, if necessary to get a wrongdoer to stop his discordant behaviour, to compensate his victims for it or to get him to reform so that he will not reoffend. However, there is probably a kind of disrespectful treatment when substantial force is used against a wrongdoer for some purpose other than getting him to ‘clean up his own mess’.

Turning now to the Western backward-looking theories, reconciliatory sentencing differs from them, and in some prima facie attractive ways. One of the most prominent objections to retributive accounts of punishment is that they fail to make sense of why a criminal justice system is worth the price. It takes a lot of time, effort, money and other resources to arrest apparent lawbreakers, to conduct a trial, to punish those who have been found guilty and to monitor their progress in a correctional setting, where merely giving people what they deserve or correcting unfairness do not seem weighty enough to justify the costs. It seems to many that a major public institution should promise to do some kind of good for society, and not merely increase the overall amount of suffering or other harm in the world in the manner of an eye for an eye. According to a reconciliatory approach, a major point of setting up and maintaining a criminal justice system includes reforming offenders so that they do not reoffend, getting them to compensate their victims and more generally healing broken relationships.

Another weakness of standard retributive theories is that they cannot easily account for intuitions that an offender’s moral reform can call for a lesser penalty. For example, it is common for judges in South Africa and elsewhere to sentence in part based on whether or not an offender has expressed remorse for having committed the crime. That should be completely irrelevant on a desert or fairness model, which directs a judge nearly exclusively to the nature of the crime committed, regardless of what has happened since. However, by a reconciliatory approach, a genuine expression of remorse could be reason to reduce a penalty, even if some kind of burden that broadly tracks the nature of the crime is essential, both to express remorse and to express disapproval. Similar remarks apply to the practice of parole, that is, early release for good behaviour. Concrete evidence of rehabilitation gives a reason to reduce a penalty in terms of a reconciliatory approach, if only marginally, but it is no reason


to reduce one according to a retributive approach, supposing the initial penalty was indeed strictly proportionate to the crime committed.

A final advantage for reconciliatory sentencing relative to Western retributivism concerns penalties that might be deserved or fair, but that are intuitively wrong to impose nonetheless. I am thinking of torture, rape and death. Torturers, rapists and murderers might well have these respective penalties coming to them, as proportionate to what they have done, but they would be intuitively unjust for a court to authorise. By an ubuntu-based reconciliation, part of the explanation of why these penalties are unjust is that they are not merely unnecessary, but also unlikely, to produce meaningful compensation for victims and moral reform on the part of offenders. In addition, these kinds of penalties are not necessary in order for offenders to express their guilt and for a court to express disapproval of their guilty behaviour, with quite weighty but less than strictly proportionate burdens being sufficient. Although these considerations probably do not constitute the entire explanation of why certain kinds of severe penalties are unjust, they are more than is available to the desert or fairness theorist.

D Some contrasts with current practices

If reconciliation were made the final end of a criminal trial, then some sentencing practices common in South Africa and in many other jurisdictions would need to be substantially revised. In particular, reconciliation would probably mean that mandatory sentences and imprisonment, the focus of *Ndlovu v The State*, would not be used as frequently as they are.

Brendan Solly Ndlovu was convicted of a particularly brutal rape and sentenced to life imprisonment. The Constitutional Court needed to decide whether the sentence was appropriate, given that Ndlovu had been charged with rape, not with the infliction of grievous bodily harm, but had been sentenced on the basis of both. The mandatory minimum sentence for a first offence of rape is 10 years in prison, with a maximum of 15, while the minimum (and, equally, maximum) for the infliction of grievous bodily harm is imprisonment for life (although parole is possible after having served 25 years). The Court ruled that the sentence of life in prison was unconstitutional, since Ndlovu had been charged only with rape. The Court instead imposed the maximum of 15 years, in accordance with the statute governing the crime for which Ndlovu had in fact been charged. Although the Court deemed criminal justice on the whole to be best served by reducing Ndlovu’s sentence, it lamented its inability to impose a much longer sentence of imprisonment on him.

Although this case concerns mandatory minimum sentences of imprisonment, these two issues are logically distinct; one could have mandatory minimums when it comes to, say fines instead of jailtime, and, then, one could imprison without a legislature having indicated which amount of time served is essential. I first argue that mandatory minimum sentences, whether of prison or some other kind of penalty, are usually unjustified and next that prison is rarely an appropriate kind of sentence.

Mandatory minimum sentences are straightforwardly justified by the Western backward-looking and forward-looking theories. If the sentences are proportionate to the nature of the crime, making allowances for mitigating and aggravating factors, then they can be deserved for having committed a certain kind of crime or be what would remove an unfair advantage obtained by having done so. And if the sentences would incapacitate or deter potential offenders to an extent that crime would be reduced to a noticeable degree, then utilitarianism and self-defence theory would also prescribe them.
In contrast, a practice of reconciliatory sentencing demands flexibility in response to the particular circumstances of offenders and victims. Above I argued that reconciliatory sentencing includes a maximum, permitting disavowal of a strength no greater than the crime, and also that it includes a minimum, in the sense of normally requiring some kind of burden to be placed on offenders so as to disavow the crime, even on those able to compensate victims and change their motivations without a burden. However, in between there would be a range of possible severities that a judge would be best placed to pick amongst, a legislature of course being uninformed about the specifics of a given case. When determining how victims should be compensated, a judge needs to know what particular ways they were harmed, what would help make up for those harms, what offenders are realistically capable of doing and what form of compensation would place appropriate burdens on offenders. Similarly, when determining how offenders should be reformed, a judge needs to know why they were motivated to offend, what would be likely to change their motivations and what would be appropriately burdensome with regard to expressive considerations. Reconciliatory sentencing requires judgement.

Somewhat similar considerations apply to imprisonment, by which I mean the predominant form where offenders are simply locked up, at best given time to think and offered some optional rehabilitative and recreational activities. Backward-looking approaches to punishment easily justify prison, since they do not require any good to come from a type of penalty. If prison is of a severity proportionate to the nature of the crime, where the severity is deserved or would correct unfairness, it is justified. And then the forward-looking theories naturally justify prison as well. Recall that, for them, incapacitation and deterrence are proper mechanisms by which punishment should be used to reduce crime, where prison renders someone unable to commit crime and tends to make prospective criminals fearful of getting caught.

Prison should not be the default mode of punishment, however, if the aim of criminal justice is reconciliation, understood as the combination of the disavowal of the wrongful discord done in the past and the improved chance of harmonious relationship in the future. Although prison can express disavowal of a crime, it does not serve the additional function of making repair of the broken relationship more likely. Merely locking someone up does not reliably foster either compensation to victims or reform of offenders.

What, then, should have been done with Ndlovu? His case is amongst the most difficult for a reconciliatory approach, given how heinous his behaviour was. Detainment would have been appropriate, but that is not necessarily the same thing as jail as we know it. Ndlovu should have been made to undertake truly burdensome reparations for his victim and to undergo difficult procedures likely to change his inclination to reoffend, both of which could have been ways of expressing remorse on his part, but at least would have been vehicles by which to express disapproval on the part of the political community.

For a start, Ndlovu of course should have apologised to his victim, a way of showing that she matters. In addition, he should have been given a way to earn money that could have been directed to her, or otherwise afforded a way to labour in ways that would have benefited her. Perhaps because of the crime she has been unable to work, and so has found it difficult to afford school fees for her children; Ndlovu could have been required to pay for them. If she did not want to be reminded of him, the court could have ordered him to help a charity of her choice or a state clinic that would offer her therapy.

Furthermore, Ndlovu should have been mandated to undergo counselling of an intense sort. With respect to his beliefs, he should have been forced to reconsider his views of the standing
of women. Perhaps he considers them to be his property or second-class citizens, and hence as something to be used as a mere means to his ends. His emotions, too, should have been explored and probably adjusted. Did he commit the brutal rape because he feels impotent and needed a sense of power? Does he hate women because of how he was reared? One hopes that, in time, his personhood would develop, so that he would feel the appropriate sort of guilt and be haunted by what he has done. And then Ndlovu also should have done what would have changed his desires. Perhaps he lacks a second-order desire to avoid desiring to rape and to inflict pain, or, if he has such a second-order desire, it might be ineffective at changing his first-order ones. Court-ordered self-exploration would have been apt, as would have been the mentoring of others less reformed, supposing there were improvement on Ndlovu’s part.

E Replies to objections

Some readers will find these penalties to be intuitively insufficient, with the prospect of compensation, reform and improved relationships not being important enough to forgo a harsher penalty such as imprisonment or even corporal punishment of some kind. However, if ubuntu is our touchstone, then we have to let go of vengeful or retributive reactions. Conversely, others will find the therapeutic interventions overly intrusive or otherwise illiberal in some way. However, the claim is not that either brainwashing or brain surgery is permissible; the methods of reform must be consistent with respect for a person’s capacity for communal relationship. Plus, a focus on the offender’s character is arguably justified, given an ubuntu ethic’s concern for personhood and supposing the development of personhood would be a particularly reliable way to prevent recidivism.

However, what is to be done if an offender refuses to undertake the burdens of compensation and reform? In that case, it might be that threats and penalties designed to prompt conformity would be appropriate, where these might involve imprisonment. Such a ‘back-up’ approach is reminiscent of that taken by South Africa’s Truth and Reconciliation Commission, where a normal trial could proceed if offenders did not fully disclose the political crimes they had committed during the apartheid era. Although not all penalties would be reconciliatory in such a scheme, it would still constitute a radical departure from the current approaches in most jurisdictions in Africa and the West.

For another objection, what if an offender is simply too dangerous to participate in compensatory or reformative procedures? In that case, confinement, in contrast to imprisonment, would be appropriate. Putting in prison, I am supposing, would involve, if not the aim to harm, at least an intervention that is likely to harm the one imprisoned. In contrast, confinement need not involve such an intention or expectation. It could be a matter of sequestering an out of control offender in a comfortable manner, similar to a quarantine. In that case, it would not count as punishment, since no hard treatment would essentially be involved and since it would not be a response to a crime that had already occurred. Preventive detention of this sort would be outside the reconciliatory approach to punishment that I have advanced here, even if it does have a small but proper role to play in a criminal justice system.

48 Interestingly, one interlocutor has suggested to me that reconciliatory sentencing, given its expressive focus, might be apt for the very worst deeds, such as crimes against humanity, but not for crimes such as theft in which symbolic considerations seem irrelevant. However, reconciliatory sentencing is apt as a way to respond to any crime in which people’s capacity to relate communally is treated disrespectfully. The disrespect in theft also calls for an expressive response in part.
Finally, one might reasonably be concerned about how realistic reconciliatory sentencing is, at least in jurisdictions in which there are many crimes but few defence attorneys, prosecutors and judges with expertise. Treating reconciliation as the end of a criminal trial would mean spending vast amounts of resources to establish guilt during the trial, to ascertain an apt sentence if guilt is established and then to carry out the sentence. In contrast, a system of plea bargaining, as standardly combined with fines or imprisonment, would be much more efficient, and could well be essential for any criminal justice system with substantial numbers of cases.

I accept the point about needing to make compromises in situations of scarce human and other resources. However, doing so is consistent with maintaining that reconciliatory sentencing is an ideal for which to strive. Reconciliatory sentencing plausibly explains not only why plea bargaining would be an unjust way of dealing with crimes in situations where there is no resource scarcity, but also how plea bargaining sacrifices moral weight even when it is, all things considered, justified. Plea bargaining streamlines criminal justice at the cost of failure to ascertain guilt with care, to hear victims out and to sentence in ways likely to compensate victims, reform offenders and adequately disavow previous wrongdoing. By my favoured account of sentencing, we should want as much of those things as we can realistically get in the present circumstances, striving for progressive realisation of more over time.

V CONCLUSION

My aim in this article has been to sketch some respects in which a judge should sentence in a criminal trial, assuming its central aim to foster an ubuntu-based conception of reconciliation. Specifically, I have advanced the view that, instead of requiring forgiveness or otherwise forbidding punishment, an attractive notion of reconciliation includes the disavowal of crime, which, in turn, typically prescribes punishment. A genuine expression of remorse on the part of offenders or disapproval on the part of the political community means placing burdens on them, in particular ones oriented towards the rehabilitation of offenders and the compensation of victims. Although this account of just punishment has grown out of characteristically African views of personhood, communion and reconciliation, it is meant to capture some intuitions that are widely shared, even by those who currently endorse more Western theories.

Supposing that reconciliatory sentencing indeed merits consideration, a number of other theoretical projects would naturally follow. For one, it is worth considering whether the evidentiary procedures of a criminal trial need to be revised so as to foster reconciliation. I have contended that they need to reveal guilt and the degree of it, but might there be a way of doing so that would be more likely, say, to prompt an apology on the part of offenders?49 What does reconciliation entail for the adversarial versus inquisitorial distinction about how to ascertain legal guilt?

For another, in this article I have set aside the question of precisely whom a crime characteristically wrongs, having focused on an individual victim, but not also her (extended) family or the broader society. However, it is well known that indigenous sub-Saharan peoples often considered legal transgressions to have a community dimension, or at least for many others beyond the ‘immediate’ or ‘direct’ perpetrator and victim to have a stake in

---

49 Cf Sachs J (note 2 above) para 117.
reconciliatory processes.\textsuperscript{50} How might a judge overseeing a criminal trial in a ‘modern’ state plausibly incorporate these facets of African criminal justice?

Beyond the responsibilities of a judge in a criminal trial, it is also worth thinking about how reconciliation might bear on those of other actors. For example, I have not systematically addressed the matter of what should be criminalised. However, viewing reconciliation to be the final end of a criminal trial probably has implications for what a legislature should count as a crime. Might it rule out, for instance, victimless activities as meriting a response from a criminal justice system? Should legislators decriminalise activities such as physician-assisted suicide and drug-taking, where these are not inherently discordant in respect of other parties, or should they instead criminalise such behaviour out of concern for citizens’ personhood?

Finally, for now, the account of sentencing given here, which includes compensation to victims as an inherent feature, raises the question of whether the distinction between criminal and civil trials should be abandoned. Normally, the sort of harm that a civil trial seeks to repair is what was caused wrongfully, suggesting that the kind of criminal trial advocated in this article would render a civil trial unnecessary. However, this inference might be too quick. Suppose, for example, that a way of compensating that would be appropriately burdensome on the offender would not provide as much repair of the harm done to the victim as some other way. Would a civil trial be apt in that situation? Or are there harms that were not wrongfully caused by a certain agent but which this agent should be forced to compensate, say, because he is best placed to do so? I submit that these and related questions merit consideration.

Unpacking Section 25: What, If Any, Are the Legal Barriers to Transformative Land Reform?

JACKIE DUGARD

ABSTRACT: The ‘land question’ has undoubtedly become the issue at the centre of South African politics. This is largely because, notwithstanding various government efforts since 1994 to redress access to land and align contested property regimes in the public interest, the white minority continues to own a disproportionally large amount of land. The untransformed pattern of ownership of (and access) to land has become a potent symbol of the broader failures of South Africa’s transition from apartheid. At the heart of the current political deliberation has been contestation over s 25 of the Constitution, the ‘property clause’, which is widely perceived to be an impediment to transformation and has been reviewed for amendment. Reflecting an increasingly uneasy bundle of imperatives from redistribution and restorative restitution, through customary land use, to private property ownership, it is unclear what any amendment will look like. At the same time as there are calls to repeal private property ownership, many poor people and communities are appealing to the government to speed up private title processes, and property owners (black and white), as well as traditional authorities, are demanding greater, rather than less, protection of their property rights. Against this backdrop, instead of interrogating what has (and has not) been achieved since 1994 or whether s 25 is optimal, this article examines the extent to which the current legal frameworks and judicial interpretation thereof are likely to constitute a hindrance to any government wishing to pursue transformative land reform. In doing so the article seeks to contribute towards a more nuanced understanding of the fault lines between the legal realm and the socio-political realm, as relating to the question of transformative land reform.

KEYWORDS: substantive transformation, section 25, land, property

AUTHOR: Associate Professor, School of Law, University of the Witwatersrand. Email: Jackie.Dugard@wits.ac.za

ACKNOWLEDGEMENTS: I am grateful to the anonymous referees for their very useful comments and to the editors for their editorial work on the article.
I INTRODUCTION

Besides the election of the new party president (and the other ‘Top Six’ officials),¹ the issue that dominated the 54th elective conference of African National Congress (‘ANC’) in Johannesburg in December 2017 was the evolving crisis around the evident failure of post-apartheid land reform.² Following what by all accounts was a heated deliberation, the ANC resolved to adopt a radical programme to fast-track land redistribution without compensation.³ Accordingly, the ANC’s ‘January 8 Statement’ of 2018 pledged ‘the expropriation of land without compensation’, albeit in a manner that ‘not only meets the constitutional requirement of redress, but also promotes economic development, agricultural production and food security’.⁴ And, following the ANC’s support of an Economic Freedom Fighters (‘EFF’) motion⁵ on 27 February 2018, the National Assembly voted to establish an ad hoc Constitutional Review Committee (‘CRC’) to investigate a possible amendment to the Constitution⁶ to allow for radical land reform.⁷ More recently, on 31 July 2018, in the midst of the CRC public hearings, President Cyril Ramaphosa announced that the ANC would amend the Constitution to enable the expropriation of land without compensation.⁸

¹ In a tight race against Nkosazana Dlamini-Zuma, Cyril Ramaphosa was elected as the 14th president of the ANC on 18 December 2017. Following the resignation of President Jacob Zuma two months later, Ramaphosa was elected President by the National Assembly.
² The term ‘land reform’ is used in this article generically to encompass all land- and property-related reform measures.
⁵ Interestingly, on 2 March 2017, the ANC voted against an EFF motion in the National Assembly to amend the Constitution to allow for the expropriation of land without compensation. At the time, then Deputy Secretary–General of the ANC, Jessie Duarte, defended the move, explaining that the Constitution allows the ANC to do what needs to be done on the land question: ‘ANC Explains its Position on Land Expropriation’ ENCA (2 March 2017) ENCA, available at https://www.enca.com/south-africa/anc-explains-its-position-on-land-expropriation.
⁸ ‘Cyril Ramaphosa: We Are Going to Amend the Constitution on Land’ Businesslive (1 August 2018), available at https://www.businesslive.co.za/rdm/politics/2018-08-01-cyril-ramaphosa-we-are-going-to-amend-the-constitution-on-land/.
Such moves, however politically motivated and as yet undefined, have occurred against the devastating reality of widening, and persistently racialised, socio-economic inequality,\(^9\) including unequal access to property, especially rural land. A legacy of South Africa’s colonial and apartheid past, which witnessed the displacement of black people on a scale that ‘far exceeded that in any other colonial state in Africa’,\(^10\) the ‘land question’ has undoubtedly become the issue at the centre of national politics. This is largely because, notwithstanding various government efforts since 1994 to redress access to land and align contested property regimes in the public interest, the white minority continues to own a disproportionately large amount of land.\(^11\) Indeed, some commentators have argued that, rather than dismantling colonial apartheid patterns, post-apartheid land reform strategies have perpetuated elite control and ownership of land ‘while the majority rural black population holds insecure land rights and is often beholden to the state [including traditional authorities] and whites for their use of the land as “partners”’.\(^12\) The untransformed pattern of ownership of (and access to) land, constituting a ‘colonial present, even in the midst of land reform’,\(^13\) has become a potent symbol of the broader failures of South Africa’s transition from apartheid. As highlighted by Elmien du Plessis, the purpose of land reform was ‘not merely to return land to claimants’, but also more comprehensively to right the wrongs of the past, to redress, to heal, to ‘reverse

racially skewed patterns of landownership’ and to ‘deracialise the privilege in property rights’, as well as to acknowledge the ‘hurt of dispossession and the histories of injustice’.14

At the heart of the current political deliberation has been contestation over s 25 of the Constitution, the ‘property clause’, which is widely perceived to be an impediment to transformation. For example, Black First Land First’s submission to the Portfolio Committee on Public Works’ public hearings on the Expropriation Bill B4-2015 (‘Expropriation Bill’) on 4 August 2015 sets out that ‘Section 25 legalises land theft and legitimises colonialism … Section 25 in its entirety is a yoke around the necks and shackles in the feet and hands of our people. It makes us slaves in our own land’.15 Officially, the National Assembly’s explicit mandate to the CRC was to consider amending s 25 to allow more radical land reform. More polemically, the Azanian People’s Organisation (‘AZAPO’) has called for the scrapping of s 25 and the nationalisation of land to (re)establish ‘black power’.16 And, academically, Magobe Bernard Ramose has argued that s 25 is a fatal obstacle to the objective of achieving justice for indigenous black South Africans.17

This consternation over the property clause is not surprising. Section 25 manages a political tension that was at the heart of South Africa’s historic constitutional negotiations: how to deal with a reality in 1994 in which whites owned the vast majority of the land, despite comprising a small minority of the population. During the negotiations the National Party and various liberal interests, along with large commercial enterprises (controlled predominantly by the white minority), on the one hand, wanted to ensure that existing ownership of property by the white minority would be protected. On the other hand, more radical wings of the ANC and aligned liberation movements saw the need for major restitution and redistribution of land to correct the historical injustices of colonialism and apartheid.18 Section 25 – as well as the precursor s 28 in the Interim Constitution of the Republic of South Africa, 1993 (Interim Constitution) – represents a political compromise between these positions, seeking to balance the rights and interests of the (overwhelmingly white) ‘haves’ against the rights and interests of the (overwhelmingly black) ‘have-nots’. Thus, s 25 undoubtedly affords a degree of protection to existing property owners by prohibiting arbitrary deprivation of property.19 However, at the same time, it includes an imperative to advance access to land on an equitable basis, and a framework to pursue land restitution, inter alia through authorising expropriation (albeit with some form of compensation) when in the public interest. The ‘public interest’ is explicitly


17 M Ramose ‘To Whom Does the Land Belong’ (2016) 50 Psychology in Society 86.


19 Constitution, s 25(1) prohibits arbitrary deprivation of property.
defined as including ‘the nation’s commitment to land reform, and to reforms to bring about equitable access to all South Africa’s natural resources’. 20

Reflecting an increasingly uneasy bundle of imperatives from redistribution and restorative restitution, through customary land use, to private property ownership, s 25’s Janus character has given rise to a schizophrenic public discourse. At the same time as there are calls to amend and even repeal s 25, many poor people and communities are appealing to the government to speed up private property titling processes, and property owners (black and white), as well as traditional authorities, are demanding greater, rather than less, protection of their property rights. Against this backdrop, instead of interrogating what has (and has not) been achieved since 1994, or whether s 25 is optimal – the socio-political question, this article examines the extent to which the current legal frameworks and judicial interpretation thereof are likely to constitute a hindrance to any government wishing to pursue transformative land reform. In doing so, the article seeks to contribute towards a more nuanced understanding of the fault lines between the legal realm and the socio-political realm, as relating to the transformative land reform question. Before unpacking s 25’s legal limits (in part III), the article first grapples in part II with the question of what transformation means in the legal realm and how this might (should?) be applied to land.

II DEFINING TRANSFORMATIVE LAND REFORM

In his influential 1998 article entitled ‘Legal Culture and Transformative Constitutionalism’, Karl Klare identifies transformative constitutionalism as the:

Long-term project of constitutional enactment, interpretation and enforcement committed (not in isolation, of course, but in historical context of conducive political developments) to transforming a country’s political and social institutions and power relationships in a democratic, participatory, and egalitarian direction. Transformative constitutionalism connotes an enterprise of inducing large-scale change through non-violent political processes grounded in law. 21

Expanding on the ‘egalitarian direction’, Klare describes transformative constitutionalism as being potentially post-liberal in its pursuit of ‘substantive equality’. 22 As explained by Cathi Albertyn, in South Africa the term substantive equality emerged ‘from the constitutional negotiations of the early 1990s’. 23 In constitutional terms, substantive equality goes beyond an ‘inclusive’ trajectory (‘inclusive equality’) that ‘would align with a liberal idea of inclusion into the status quo’. 24 Inclusive equality is referred to by Nancy Fraser as an ‘affirmative’ approach to change, in which the objective is to correct ‘inequitable outcomes of social arrangements without disturbing the underlying framework that generates them’. 25

20 Constitution, s 25(2) provides that property may be expropriated only in terms of a law of general application ‘for a public purpose or in the public interest’ and subject to compensation; s 25(4) specifies that ‘the public interest’ includes ‘the nation’s commitment to land reform’; and s 25(2)–(9) sets out the parameters of expropriation and land restitution. Land restitution is expounded in greater detail in the Restitution of Land Rights Act 22 of 1994.
22 Ibid at 151.
24 Ibid at 256.
Contradistinctively, and in Klare’s post-liberal iteration, substantive equality entails a commitment to the eradication of systemic inequalities and to shifting ‘the power relations that maintain the status quo’. That substantive equality in the socio-economic (including land/property) realm would have to entail significant redistribution (especially in profoundly unequal societies such as South Africa) is emphasised by Fraser. Importantly, a constitutional commitment to transformation defined in substantive equality terms – ‘substantive transformation’ – is both an aspirational ideal and a presumption that this ‘is (at least partly) possible through law’. However, 25 years after South Africa’s transition to democratic rule, at the political level it remains unclear which model of equality we are pursuing as our vehicle for, and vision of, transformation. To an extent, as explored below, this lack of political direction has been reflected in constitutional adjudication, which still has to develop a substantive theory of socio-economic transformation.

Until now, the political trajectory – as pursued, however incompletely, through policies such as Black Economic Empowerment – has arguably been more closely aligned with inclusive than substantive equality (mainly representing an attempt to deracialise the existing unequal socio-economic order). Yet, during 2017, politicians from an increasingly embattled ANC government began to call publicly for radical economic transformation. Recently, these calls have focused on land and, by extension, the property clause. Reflecting the broader opaqueness about what kind of transformation to pursue, criticisms of s 25 reveal several strands that delineate critical societal fractures.

One strand of critique holds that s 25 imposes a western notion of property rights, and needs to be ‘decolonised’. While it is true that s 25 recognises individual private property ownership (but does not establish a positive right to this), it also implicitly (and explicitly through related legislation) encompasses plural land rights, including African customary land rights. Yet, this inclusion of customary land rights presents a political and constitutional conundrum. In geographic terms, the areas in which customary land rights apply are the previous ‘bantustans’ or ‘homeland’ (black African) areas, which are still governed by traditional authorities who exercise constitutionally-enshrined chiefly authority over customary land rights. With traditional

---

26 Albertyn (note 23 above) at 256.
27 Fraser (note 25 above) at 68–93.
28 Albertyn (note 23 above) at 254.
31 That dissatisfaction over the failure of land reform has focused on the legal sphere, as opposed to the political sphere, makes sense from the ANC’s perspective, as the political party that has governed since 1994. It is less clear why the EFF, which was established in opposition to the ANC, has pursued such a legal focus.
32 This article does not attempt an interrogation of any of the strands of academic or popular critique.
34 Constitution, s 25(6) requires the government to enact legislation to secure land tenure (as opposed to ownership) rights. As examined in part III below, a range of legislation has been enacted to protect land tenure rights, including legislation to protect customary land rights. In addition, Constitution, s 211(3) underscores that customary law is a recognised source of law in South Africa.
35 Traditional leadership is recognised as one of the institutions of government, and its powers are consolidated in ss 211 and 212 of the Constitution as well as the Traditional Leadership and Governance Framework Act 41 of 2003.
leadership typically assuming crucial decision-making powers over land, individuals and communities in customary land rights areas often operate as ‘subjects’ rather than ‘citizens’, when compared with the greater degree of social and political (if not necessarily economic) autonomy and accountability afforded in non-customary land rights areas.

This bifurcated system of land governance has resulted in tangible discontent and tension within customary communities, as poignantly detailed in the November 2017 Report of the High Level Panel on the Assessment of Key Legislation and the Acceleration of Fundamental Change (‘HLPR’). Indeed, customary communities have actively resisted the attempts by the Jacob Zuma Presidency to entrench chiefly power over land rights. For example, the Communal Land Rights Act 11 of 2004 (‘CLARA’), which sought to entrench the powers of traditional authorities to control the occupation, use and administration of customary communal land, was (successfully) challenged by rural communities in the case of Tongoane, resulting in the striking down of CLARA. Similarly, various iterations of the Traditional Courts Bill (‘TCB’) have had to be abandoned in the face of resistance among affected communities towards the extensive powers envisaged for traditional courts. And the attempt to re-open the land restitution process via the Restitution of Land Rights Amendment Act 15 of 2014 (‘RLRAA’).

Mahmoud Mamdani’s seminal 1996 book, Citizen and Subject: Contemporary Africa and the Legacy of Late Colonialism, outlines how colonial rule in Africa took on the form of a bifurcated state, with racial domination mediated through direct rule in urban areas and through tribally organised traditional authorities elsewhere, and how this bifurcation has persisted in the post-colonial period creating ‘citizens’ in urban areas and maintaining ‘subjects’ under authoritarian traditional rule in rural areas. For an analysis of the continuity in South Africa of traditional leadership through colonialism and apartheid until today, see M Buthelezi and D Skosana The Salience of Chiefs in Postapartheid South Africa: Reflections on the Nhlapo Commission (2018) in J Comaroff & J Comaroff (eds) The Politics of Custom: Chiefship, Capital, and the State in Contemporary Africa 110. In addition to diminished citizenship, scholars have pointed to the institutional perpetuation of patriarchy under traditional authorities, including regarding allocation of land (A Claassens ‘Recent Changes in Women’s Land Rights and Contested Customary Law in South Africa’ (2013) 13 Journal of Agrarian Change 71). Other scholars have highlighted the mounting fracture between customary communities and traditional leadership over mineral rights (S Mnwana “Custom” and Fractured “Community”: Mining, Property Disputes and Law on the Platinum Belt, South Africa (2016) 1 Third World Thematics: A TWQ Journal 218).


T Thipe & Others v National Minister for Agriculture and Land Affairs & Others 2010 [ZACC] 10, 2010 (6) SA 214 (CC)(‘Tongoane’).

was derailed after it was successfully challenged in the *Land Access Movement*\(^{41}\) case by affected communities, who feared that the re-opening of the restitution process would mean that their as yet unresolved claims might never be settled. Commentators have pointed out that the RLRAA was viewed by former President Zuma as a move to appease traditional leaders following the failures of CLARA and the TCB.\(^{42}\) At the time, then President Zuma had been mobilising traditional authorities ‘to hire good lawyers and make vast claims on behalf of their people’, which many rural communities understood to be a push to ‘consolidate chiefly power at the expense of the rights and citizenship’ of customary communities.\(^{43}\)

Returning to the legal realm, while the HLPR has exposed the dissatisfaction among customary communities with their current land rights governance regime as overseen by traditional authorities, it remains unclear precisely what land rights regime should be pursued in customary areas (whether communal land use, communal property ownership, private land ownership etc.)\(^{44}\) as well as how to ensure democratic citizenship over such land rights under traditional authority administration of land. The recent calls for land expropriation without compensation, including those that stress the link between restoration of land and recognition of African culture, have mostly been silent about how to approach the issue of traditional authority power over customary land rights. Interestingly, support expressed by the ANC at its elective conference in December 2017 for the HLPR’s recommendation to amend or repeal the KwaZulu Ingonyama Trust Act 3KZ of 1993 (Ingonyama Trust Act\(^{45}\) dissolved in March 2018 following strident opposition by King Goodwill Zwelithini, who warned the government to leave the Ingonyama Trust Act alone or ‘face something else’\(^{47}\).

Another (not unrelated) strand of critique focuses on the abolishment of private ownership of land as it currently operates. This strand is exemplified by the EFF’s official policy position that ‘all land should be transferred to the ownership and custodianship of the state’ and that ‘once the state is in control and custodianship of all land, those who are currently using the land or intend using the land … will apply for land-use licenses, which should be granted only

---

\(^{41}\) *Land Access Movement of South Africa & Others v Chairperson of the National Council of Provinces & Others 2016 (ZACC) 22, 2016 (5) SA 635 (CC)(‘*Land Access Movement’)*.

\(^{42}\) Centre for Law and Society: Rural Women’s Action Research Programme ‘Restitution of Land Rights Amendment Bill’ (January 2014), available at https://open.uct.ac.za/handle/11427/2497.

\(^{43}\) N Gasa ‘Laws for Traditional Leaders Emulate the Logic of Apartheid’ (14 October 2014) *Business Day* (Referring to then President Zuma’s opening address to the House of Traditional Leaders on 27 February 2014). See also A Claassens ‘Haste over Land Rights Bill not just in Aid of Buying Votes’ (10 April 2014) *Business Day* (who notes that, less than a week after President Zuma’s opening address, King Goodwill Zwelithini announced that he had agreed to submit one consolidated claim for all the land taken from the Zulu nation, and that such land would be distributed ‘fairly’ among communities once restituted).

\(^{44}\) In the absence of appropriate and acceptable legislation since the overturning of CLARA, the Interim Protection of Informal Land Rights Act 31 of 1996 (‘IPILRA’) (which was meant to be an interim piece of legislation lasting only until 31 December 1997), has had to be renewed each year since 1995). Among the recommendations of the HLPR is that urgent attention should be given to consolidating the legislation governing customary land rights (HLPR (note 37 above) at 269).

\(^{45}\) The Ingonyama Trust Act was a last-minute political compromise that was enacted just days before South Africa’s historic election on 27 April 1994. Controversially, it vests sole control over affected land with the Ingonyama Trust Board.

\(^{46}\) HLPR (note 37 above) at 273–274.

when there is a purpose for the land being applied for’.48 This would certainly be one way to transform land rights in South Africa. In relation to this article’s inquiry, it is likely that there would be significant constitutional and statutory obstacles to this kind of transformation of property rights. Indeed, this magnitude of change (that might be called ‘revolutionary equality’ or ‘revolutionary transformation’ as it would completely undo, as opposed to redistribute, ownership of land) would require an entirely different socio-political and legal order that goes beyond the scope of this article’s analysis. However, it is unclear whether, for better or worse, any political party (including the EFF) would actually (as opposed to rhetorically) abolish private title outright.49 The extent to which any such political stance would garner popular support is furthermore unclear. Case studies suggest that low- and no-income households wish for more, rather than less, property ownership.50 And, presumably, the majority of black property owners (including beneficiaries of the state’s subsidy housing scheme51), like their white counterparts, are unlikely to support any such agenda. It should also be noted that, alongside property ownership (which in South Africa includes public property ownership, private property ownership and communal property ownership), s 25 covers other property rights – namely, use (including customary use) and occupation (including unlawful occupation). Nonetheless, as explored in part III below, it is inescapable that, without concerted political efforts, where title (whether private, public or communal) exists alongside other property rights, ownership rights will almost always dominate, not only in the interests of the empowered but also in the aspirations of the disempowered.52

Regrettably, the Parliamentary CRC hearings that took place between June and August 2018 did not focus on the significant unanswered questions about what forms of land and

48 The EFF describes its policy position on land as being similar to the licensing scheme in the Minerals and Petroleum Resources Development Act 28 of 2002 (‘MPRDA’) in terms of which old order mineral rights were converted – with equity- and social justice-related conditions – into new order ‘transformed’ mineral rights. In Agri South Africa v Minister for Minerals and Energy 2013 [ZACC] 9, 2013 (4) SA 1 (CC) (‘AgriSA’), the Court found that the MPRDA’s licensing scheme was constitutional and did not amount to uncompensated expropriation (see discussion of AgriSA in part III below). However, the constitutionality of the MPRDA scheme relies in part on the fact that there was already a regulatory scheme in place for the licensing of mining rights, and the MPRDA constituted merely the latest in a long line of regulatory changes to the scheme governing mineral rights. Since there is no such existing regulatory framework for the licensing of property rights (the requirement for the registration of relevant property rights is not comparable), the kind of change that would be entailed in converting to the EFF’s proposed scheme for land is likely to pose a full-frontal challenge to s 25.

49 The EFF has been inconsistent in its statements about property. See for example the news report by E Gerber ‘EFF on Land Expropriation: “No One Will Lose Their House”’ (27 February 2018) News24, available at https://www.news24.com/SouthAfrica/News/eff-on-land-expropriation-no-one-will-lose-their-house-20180227. See also Kepe & Hall (note 12 above) at 129) (who point to a possibly relevant distinction between the EFF’s election song ‘Izwe Lami’ (the Land is Mine) and the Pan Africanist Congress (PAC)’s slogan from the 1960s – ‘Izwe Lethu’ (the Land is Ours)).


51 In research I undertook in April 2018 among 21 beneficiaries of the government’s private housing programme (so-called ‘RDP’ or ‘Breaking New Ground’ housing), I asked home-owners whether, if they were given the hypothetical choice between their private title RDP house (which they have to maintain) or free housing that the municipality would maintain, 20 of the 21 respondents answered that they would choose the RDP house option (Dugard ‘Staircase or Safety Net: Examining the Socio-Economic Meaning of RDP Home Ownership to Beneficiaries – a Case Study of Klapmuts, Stellenbosch (forthcoming)).

52 Dugard & Ngwenya (note 50 above).
property rights South Africans (particularly black South Africans) want, and how individuals and communities envisage land reform. In the absence of clear answers to these questions – and assuming that, for a significant proportion of black South Africans, the burning issue is not the outright abolishment of property ownership (revolutionary equality) but rather how to gain more and/or stronger property rights, in whatever form – the remainder of this article explores the extent to which the current legal parameters (in terms of both legislative text and judicial interpretation thereof) enable transformation of the land regime in the substantive equality sense.

III UNPACKING SECTION 25

Section 25 contains eight sub-sections, making the property clause the longest in the Bill of Rights.\(^{53}\) Sub-sections 25(1), (2) and (3) constitute the more defensive sub-sections, with s 25(4)–(8) constituting the more reformist sub-sections.\(^{54}\) In 1996, at the time the Final Constitution came into effect, it was clear that the defensive sub-sections would overwhelmingly act to the benefit of white, historically advantaged persons (who constituted the vast majority of property owners); whereas the reformist sub-sections were focused on benefiting historically disadvantaged black persons. It is perhaps worth noting that, over time and, as more black people have become property owners, the defensive sub-sections serve all property owners, while the reformist sub-sections continue to benefit historically disadvantaged black persons.

To be sure, one of the arguments for including property rights’ protections after 1994 was to extend these to the previously excluded black majority, whose property-related rights had been regularly disrespected and violated under apartheid. As voiced by Constitutional Court Judge, Albie Sachs, in *Port Elizabeth Municipality*:

>The blatant disregard manifested by racist statutes for property rights in the past makes it all the more important that property rights be fully respected in the new dispensation, both by the state and private persons. Yet such rights have to be understood in the context of the need for the orderly opening-up or restoration of secure property rights for those denied access to or deprived of them in the past.\(^{55}\)

Naturally, particularly from a Marxist perspective,\(^{56}\) the increasingly (but still far from representative) multi-racial profile of, specifically, private ownership rights\(^{57}\) is not evidence of any inherently transformative (or non-transformative) character of s 25. Rather, the legal

\(^{53}\) Mogobe Bernard Ramose notes that the ‘longest clause in constitutions is normally the clause that has to do with the abrogation of rights’ (note 17 above) at 97. However, in the case of s 25 of the South African Constitution, the reformist clauses that advance rights to historically disadvantaged individuals and groups are longer than the protective clauses.


\(^{55}\) *Port Elizabeth Municipality v Various Occupiers* 2004 [ZACC] 7, 2005 (1) SA 217 (CC) (‘*Port Elizabeth Municipality*’) para 15.

\(^{56}\) There has been surprisingly scant Marxist/class-based critique of the current land debates. Most commentary has been in the form of a race- and class-blind defence of property ownership rights from interest groups such as the Free Market Foundation; or a (class-de-emphasised) race-based critique of the property paradigm from scholars in the decoloniality or Black Consciousness traditions.

\(^{57}\) Statistics for private house ownership (as opposed to ownership of farm land) indicate that a higher proportion of African households (approximately 55 per cent) own their houses outright than do the proportion of white households (approximately 40 per cent) (South African Institute of Race Relations *South Africa Survey 2017* (2017) 706–707). This is due to the government’s subsidy housing programme, which results in beneficiaries becoming owners without having to make any mortgage bond payments.
potential and legal limits of s 25’s clauses, along with the associated legislation, are revealed through an examination of the textual parameters and judicial adjudication thereof.

A Section 25(1): arbitrary deprivation

Section 25 does not provide a positive guarantee of continued ownership, or a right to property per se.\(^{58}\) Rather, s 25(1) establishes a negative right not to be arbitrarily deprived of property, stating that no-one ‘may be deprived of property except in terms of a law of general application’, and ‘no law may permit arbitrary deprivation of property’. This provision protects property owners, whether historically advantaged or historically disadvantaged, from being arbitrarily deprived of their property.\(^{59}\) Although on the surface this is a conservative right (preserving property ownership), the courts have interpreted this sub-section transformatively in two key respects.

First, in terms of the relationship between deprivation and expropriation (with expropriation constituting the harshest and most comprehensive form of deprivation),\(^{60}\) the courts have on the whole interpreted deprivation quite widely and, concomitantly, interpreted expropriation quite narrowly.\(^{61}\) This is a progressive approach because, in light of the requirement for compensation in the context of expropriation, a more restrictive interpretation of what constitutes expropriation (and a wider interpretation of deprivation) means having to draw less on public funds to compensate property owners.

Similarly, the courts have clarified that, while wholesale regulatory regime-change projects embarked on by the post-apartheid state may amount to (non-arbitrary) deprivation, these do not amount to expropriation, relieving the state of the obligation to pay compensation to

---

\(^{58}\) Section 25’s precursor, s 28 of the Interim Constitution, did not contain a conventional property clause either but it had a more robustly articulated protection of property rights in that s 28(1) recognised the right of everyone to ‘acquire and hold rights in property’. This formulation was dropped in favour of the weaker iteration of property rights pursued in s 25 of the Final Constitution. In addition, in relation to arbitrary deprivation, s 28(2) of the Interim Constitution provided a stronger protection against deprivation than that of s 25(1) of the Final Constitution, stating: ‘No deprivation of any rights in property shall be permitted otherwise than in accordance with a law’ (Pienaar (note 54 above) at 167–191).

\(^{59}\) Andre van der Walt explains that a deprivation (such as the inability to transfer one’s immovable property without obtaining a municipal rates clearance certificate, or having to accommodate unlawful occupiers while the state provides alternative accommodation) restricts or limits the use of private property in the public interest without necessarily taking away the property – it affects everyone in that situation more or less equally, whereas expropriation entails taking away the property from specific owner(s) for a public use (A Van der Walt & G Pienaar Introduction to the Law of Property (6th Ed, 2013) 313).


\(^{61}\) A disappointing exception is the Constitutional Court’s decision in Arun Property Development (Pty) Ltd v City of Cape Town [2014] ZACC 37, 2015 (2) SA 584 (CC) (‘Arun’), in which, instead of finding that the City of Cape Town’s unauthorized, excessive vesting of property was simply unlawful administrative action to be set aside, the Court found this action to have amounted to expropriation and that the property developer was therefore entitled to compensation. Arun’s legal reasoning has been questioned and, arguably, does not reflect a clear legal line in relation to the jurisprudence on deprivation and/or expropriation. See J Dugard & N Seme ‘Property Rights in Court: An Examination of Judicial Attempts to Settle Section 25’s Balancing Act re Restitution and Expropriation’ (2018) 34 South African Journal on Human Rights 33, 51–55. See also E Marais ‘A Common-Law Presumption, Statutory Interpretation and Section 25(2) of the Constitution – A Tale of Three Fallacies: A Critical Analysis of the Constitutional Court’s Arun Judgment’ (2016) 133 South African Law Journal 629.
the multitude of affected property owners. In AgriSA, the Constitutional Court (the Court) endorsed the transformative objective of the Minerals and Petroleum Resources Development Act 28 of 2002 (‘MPRDA’), which seeks to ‘promote equitable access to the nation’s mineral and petroleum resources to all the people of South Africa and to ‘substantially and meaningfully expand opportunities for historically disadvantaged persons, including women, to enter the mineral and petroleum industries and to benefit from the exploitation of the nation’s mineral and petroleum resources’. Crucially, AgriSA validated the MPRDA’s conversion scheme, which had the effect of terminating the rights of any old-order mineral rights holder that failed to convert its rights to new-order mineral rights in terms of the process set out in the MPRDA. The Court found that, although constituting (non-arbitrary) deprivation, the scheme did not amount to expropriation and, as such, was not subject to compensation. The judgment thus confirmed the MPRDA’s transformative objectives, as well as the project to redistribute mineral rights from old-order (overwhelmingly historically advantaged) to new-order (historically disadvantaged) holders.

Second, the courts have clarified that property owners might have to tolerate some degree of deprivation of their property rights in the context of balancing property rights with other constitutional rights such as unlawful occupiers’ housing rights (set out in section 26 of the Constitution), and that having to accommodate unlawful occupation for months at a time will not in itself constitute arbitrary deprivation. In Blue Moonlight, which concerned an application by a private property developer to evict 86 desperately poor unlawful occupiers from an inner city building, the Court ruled that, although the property owner could not be expected to be burdened with providing accommodation to the occupiers indefinitely, ‘a degree of patience should be reasonably expected of it’ while the municipality lined up alternative accommodation. In the Blue Moonlight case, the Court operationalised this principle by ordering the municipality to provide the occupiers with alternative accommodation within four months of the judgment being delivered, and authorising an eviction only after the occupiers had been relocated in the alternative accommodation.

As underscored by the Court, the property owner, who ‘wishes to exercise its right to develop its property’, cannot be expected indefinitely to provide free housing to the occupiers, but its rights as property owner must be interpreted within the context of the requirement that eviction must be just and equitable:

The South African constitutional order recognises the social and historical context of property and related rights. The protection against arbitrary deprivation of property in section 25 of the

---

62 AgriSA (note 48 above).
63 MPRDA, s 2.
64 Constitution, s 26 (‘[E]veryone has the right to have access to adequate housing’; the ‘state must take reasonable legislative and other measures, within its available resources, to achieve the progressive realisation of this right’; and no-one ‘may be evicted from their home, or have their home demolished, without an order of court made after considering all the relevant circumstances’ and no ‘legislation may permit arbitrary evictions’).
65 City of Johannesburg Metropolitan Municipality v Blue Moonlight Properties 39 (Pty) Ltd & Another [2011] ZACC 33, 2012 (2) SA 104 (CC)(‘Blue Moonlight’).
66 Ibid at para 101.
67 Ibid para 104.
68 Ibid para 3.
69 Ibid para 97.
Constitution is balanced by the right of access to adequate housing in section 26(1) and the right not to be evicted arbitrarily from one’s home in section 26(3).70

Historical context is relevant to one’s understanding of the constitutional protection against arbitrary deprivation of property and to access to adequate housing. Apartheid legislation undermined both the right of access to adequate housing and the right to property. Section 25 prohibits arbitrary deprivation of property but also addresses the need to redress the grossly unequal social conditions … Unlawful occupation results in a deprivation of property under section 25(1). Deprivation might however pass constitutional muster by virtue of being mandated by a law of general application and if not arbitrary. Therefore [the Prevention of Illegal Eviction from and Unlawful Occupation of Land Act 19 of 1998] allows for eviction of unlawful occupiers only when it is just and equitable … A court must consider an open list of factors in the determination of what is just and equitable. The relevant factors to be taken into account in this case are the following. The Occupiers have been in occupation for more than six months. Some of them have occupied the property for a long time. The occupation was once lawful. Blue Moonlight was aware of the Occupiers when it bought the property. Eviction of the Occupiers will render them homeless. There is no competing risk of homelessness on the part of Blue Moonlight, as there might be in circumstances where eviction is sought to enable a family to move into a home. It could reasonably be expected that when land is purchased for commercial purposes the owner, who is aware of the presence of occupiers over a long time, must consider the possibility of having to endure the occupation for some time. Of course a property owner cannot be expected to provide free housing for the homeless on its property for an indefinite period. But in certain circumstances an owner may have to be somewhat patient, and accept that the right to occupation may be temporarily restricted, as Blue Moonlight’s situation in this case has already illustrated. An owner’s right to use and enjoy property at common law can be limited in the process of the justice and equity enquiry mandated by [the Prevention of Illegal Eviction from and Unlawful Occupation of Land Act 19 of 1998].71

Like the hierarchical contestation between traditional authorities and customary communities outlined in part II above, the clash between private ownership and unlawful occupation rights is an important structural fault line in the property paradigm. While not specifying how this clash of rights should be governed, s 25 provides a framework in which maximum latitude is afforded to the executive, legislature and judiciary to pursue a substantive equality approach. The extent to which, in the absence of concerted executive action, the judiciary has adjudicated such clashes transformatively is explored in the analysis of s 25(6) in part C below.

B  Section 25(2), (3) and (4): expropriation for a public purpose or in the public interest72

The power of the state to expropriate private property – sometimes referred to as eminent domain – is a common feature of most legal frameworks and is regularly used by governments around the world to pursue the public purpose of building roads or dams etc. In South Africa, expropriation generally is set out in s 25(2), which establishes that property can be expropriated only in terms of a law of general application–

a. for a public purpose or in the public interest; and

b. subject to compensation, the amount of which and the time and manner of payment of which have either been agreed to by those affected or decided or approved by a court.

70  Ibid para 34 (footnotes omitted).
71  Ibid paras 35–40 (footnotes omitted).
Apart from the usual public purposes, s 25(4) explicitly authorises the state to expropriate ‘in the public interest’\(^{73}\) for the purposes of land reform, land restitution and ‘to bring about equitable access to natural resources’, and clarifies that ‘property is not limited to land’.\(^{74}\) Nonetheless, s 25’s expropriation framework has been popularly cast as inimical to transformation because of its supposed requirement of a ‘willing buyer, willing seller’, market value-driven compensation approach. This approach, which for largely undefended reasons has to date been pursued by the government, is in fact not mandated by the Constitution.

It is probable that the persistence of a ‘willing buyer, willing seller’, market value-driven compensation approach in the post-apartheid era is at least partially explained by a mistaken continued reliance on s 12(a)(i) of the Expropriation Act 63 of 1975 (Expropriation Act), which refers to compensation for expropriation reflecting ‘the amount which the property would have realised if sold on the date of notice in the open market’. Under the apartheid regime, this clause certainly did mandate a ‘willing buyer, willing seller’ market value-driven approach. Crucially, however, the Expropriation Act must comply with the Constitution to be lawful. This means that s 25(3) of the Constitution’s formulation for compensation must take precedence.\(^{75}\) Section 25(3) provides:

> The amount of the compensation and the time and manner of payment must be just and equitable, reflecting an equitable balance between the public interest and the interests of those affected, having regard to all relevant circumstances, including—
> a. the current use of the property;
> b. the history of the acquisition of the property and use of the property;
> c. the market value of the property;
> d. the extent of direct state investment and subsidy in the acquisition and beneficial capital improvement of the property; and
> e. the purpose of the expropriation.

It is clear from s 25(3)(a)–(e) that market value is simply one of a range of (non-exhaustive) factors to be considered when deciding how much compensation to award in cases of

---

\(^{73}\) Pienaar notes that s 28 of the Interim Constitution had a ‘rather limited scope’ for expropriation, allowing expropriation only for a ‘public purpose’, which was reformulated for the (Final) Constitution specifically to acknowledge the need for land reform and restitution (Pienaar (note 54 above) 170).

\(^{74}\) The provisions in s 25(4) that the ‘public interest’ includes the ‘nation’s commitment … to reforms to bring about equitable access to all South Africa’s natural resources’ and that ‘property is not limited to land’, create significant room for the state to undertake systemic transformation of property regimes such as water, minerals, land etc. There is not the space here to deal with the resources aspect, suffice to mention that there has been a wholesale restructuring in relation to water ownership and rights, overseen by the Water Services Act 108 of 1997 and the National Water Act 36 of 1998 and, as outlined above, a similar substantial restructuring of the regime governing mineral rights under the MPRDA.

\(^{75}\) While the subsidiarity principle establishes that where there is legislation to give effect to a constitutional right, any person alleging a violation of the right should rely on the legislation rather than directly on the Constitution, direct reliance on the Constitution is possible when the legislation does not properly give effect to the constitutional right (Pienaar (note 54 above) 188). Nonetheless, a cleaner approach would be for the Expropriation Act to be amended to bring it in line with the Constitution. Parliament has been attempting to enact a new Expropriation Act for many years. The latest version of the Expropriation Bill was approved by Parliament on 26 May 2016. However, in mid-2016 the Bill was sent back to Parliament by then President Zuma over concerns related to the public participation process pursued by the National Council of Provinces. Intriguingly, on 28 August 2018, the Public Works Committee withdrew the Expropriation Bill (L Ensor ‘Expropriation Bill Withdrawn for Further Consideration’ (28 August 2019) Business Live, available at https://www.businesslive.co.za/bd/national/2018-08-28-expropriation-bill-withdrawn-for-further-consideration/).
expropriation. This list implies a system of calculation in which even if market value is used as the starting point of any calculation of compensation, after proper consideration of all the other factors, the final amount could be substantially lower. A recent Land Claims Court (‘LCC’) case, Msiza LCC – concerning s 16 of the Labour Tenants (Land Reform) Act 3 of 1996 (‘LTA’), which enables labour tenants to acquire ownership of the land on which they have worked and lived – confirmed this legal interpretation. Msiza LCC emphasised that the guiding principle in s 25(3) is for just and equitable compensation rather than market value compensation, meaning that, even if market value is a useful starting point in deciding the amount of compensation (the ‘two-stage’ approach to calculation of compensation), a court can award below-market value compensation in the public interest:

In this trial, the third and fourth respondents [the affected property owners] were insistent upon the payment of market value for compensation. I must dispense with this argument at this early stage. Market value is not the basis for the determination of compensation under s 25 of the Constitution where property or land has been acquired by the state in a compulsory fashion. The departure point for the determination of compensation is justice and equity. Market value is simply one of the considerations to be borne in mind when a court assesses just and equitable compensation. It is not correct to submit, as was done on behalf of the landowners, that the jurisprudence of this court has installed market value as a pre-eminent consideration. Properly understood, the jurisprudence of this court shows that market value is regularly used as an entry point to the analysis because it is the most tangible factor in all of the factors listed in s 25(3). This is not to make market value the most important factor in the analysis of just and equitable compensation; the object is always to determine compensation which is just and equitable, not to determine the market value of the property.

Although the Supreme Court of Appeal (‘SCA’) subsequently overturned the quantum of compensation awarded in Msiza LCC on a (questionable) technicality, the legal principle regarding the relative weight of the market value criterion is well-established. For example, in the prior expropriation case of Khumalo, the LCC used market value as merely the starting point for its determination of the value of compensation. And in Du Toit, the Court...

76 So, for example, according to the formulation of s 25(3), where the property had been egregiously dispossessed, was not being used for food crops, had benefitted from substantial state subsidies under apartheid, and where the expropriation was going to result in the restitution of the land to a community of farmers, compensation might be extremely low. That s 25(3)’s matrix for calculating compensation to current owners has not typically been applied as described here is because – arising from the prevailing administrative practice of pursuing a market value sale approach – there have been hardly any instances in which expropriation has been used in land reform cases. However, should the government decide to pursue expropriation and to calculate compensation along these lines, s 25 provides an enabling framework to do so.

77 Msiza v Director-General, Department of Rural Development and Land Reform & Others [2016] ZALCC 12, 2016 (5) SA 513 (LCC) (‘Msiza LCC’).
78 Ibid paras 29–30.
79 In September 2017, the Msiza LCC judgment was successfully appealed to the SCA, which, while not disputing the framework for the calculation of compensation of the LCC judgment, disagreed with the calculation and awarded a higher amount to the landowners (Uys NO & Another v Msiza & Others [2017] ZASCA 130, 2018 (3) SA 440 (SCA)). Regrettably, the Department of Rural Development and Land Reform failed to seek leave to appeal the judgment to the Constitutional Court (where, arguably, the Court would have endorsed the LCC rather than the SCA interpretation and calculation).
emphasised the Constitution’s requirement that compensation be ‘just and equitable’, having regard to ‘all relevant circumstances’ including (but not limited to) those listed in s 25(3).81

Beyond market value, the other factors included in the mandatory but non-exhaustive list in s 25(3) have a transformative, public interest basis. The current use of the property is a consideration that establishes justification for the expropriation of scarce resources, such as land and minerals, where these are ‘not being used productively and they are needed for reformatory purposes, such as access to housing or access to the mining industry for historically disadvantaged parties, or to support emerging farmers’.82 The history of the acquisition of the property implicitly authorises a discounting of the amount of compensation in respect of any property acquired as a result of forced removals and/or made available to white farmers at discounted prices and/or accompanied by state subsidies.83 This is linked to the consideration of ‘the extent of direct state investment and subsidy’, referencing the inequity of current owners who have received substantial state subsidies over the years receiving market value compensation (amounting to a double subsidy for historically advantaged persons) in instances of expropriation. Finally, ‘the purpose of the expropriation’ offers justification ‘for expropriations that are aimed at alleviating pressing social needs’, and also justifying ‘downward adjustment of the amount of compensation’.84

It is true that the reference in s 25(2)(b) to any such expropriation being ‘subject to compensation’ suggests that some form of compensation must be awarded. However, it is possible that on a proper assessment of all the factors set out above, the compensation could be very low and even nothing at all (this argument is taken further in the analysis of s 25(8) below). Moreover, regarding the requirement in relation to compensation that ‘the amount of which and the time and manner of payment of which have either been agreed to by those affected or decided or approved by a court’, it has been conclusively established by the Court in Haffejee that, while it is ideal for the amount, time and manner of compensation to be established prior to the expropriation, this is not necessary.85 In other words, an owner may not hold up an impending expropriation by arguing over the price.

From the above it is clear that s 25 mandates neither a willing buyer/willing seller nor market value compensation regime for expropriation. As stressed by the LCC in both Khumalo and Msiza, and the Court in Du Toit, the constitutional framework of ‘just and equitable compensation’ (rather than the Expropriations Act) should govern expropriations. The courts could arguably have done more to establish a substantively transformative approach for applying s 25(3)’s compensation matrix. But their ability to do so has been circumscribed by the perplexing adherence of the post-apartheid government to a market-value, willing buyer, willing seller paradigm, which has meant that the courts have had limited opportunities to adjudicate this issue in the context of land reform. To the extent that the law per se frustrates any substantive transformation in relation to the calculation of compensation for current (historically advantaged) owners, it is the Expropriation Act’s contradiction of s 25’s paradigm. Therefore, if the CRC is looking for legislation to amend, the Expropriation Act would be a

---

83 Ibid at 129.
84 Ibid.

---
good place to start. It is possible that the withdrawal of the Expropriation Bill on 28 August 2018 signals that such a process is underway.

C Section 25(5): foster conditions to gain access to land, and section 25(6): advance tenure security

Sections 25(5) and 25(6) are widely permissive authorisations for the state to pursue programmes to advance access to land and tenure security. Advancing access to land could be in the form of facilitating land ownership or it could entail making land available for secure occupation and use. Regarding land ownership, s 25(5) clearly empowers – and indeed obliges – the state to pursue land redistribution (including via expropriation). However, as signalled by the HLPR, the scope of s 25(5) of the Constitution has not yet been interpreted judicially, so there is no jurisprudence on what constitutes adequate measures to ‘enable citizens to gain access to land on an equitable basis’. Nonetheless, to give legal effect to this mandate, two of the main laws that have been enacted are the LTA and the Provision of Land and Assistance Act 126 of 1993 (‘PLAA’).

The LTA exists to benefit a particular category of land user: a person who resides on, or has the right to reside on, the farm or another farm of the owner; who has or has had the right to use cropping or grazing land on the farm in exchange for labour; and whose parent or grandparent resided or resides on such farm or had the use of cropping or grazing land in exchange for labour. The redistributive objective of the LTA is contained in ch III, which sets out parameters for the acquisition of ownership by labour tenants. Section 16(1) of the LTA establishes that a labour tenant can apply for an award of land ‘which he or she is entitled to occupy or use in terms of section 3’, which provides that ‘a person who was a labour tenant on 2 June 1995 shall have the right with his or her family members to occupy and use’ that part of the farm which ‘he or she or his or her associate was using and occupying on that date’; or ‘the land which he or she or his or her family occupied or used during a period of five years immediately prior to the commencement of this Act, or of which he or she or his or her family was deprived contrary to the terms of an agreement between the parties’; or ‘rights in land elsewhere on the farm or in the vicinity which may have been proposed by the owner of the farm’ – any such claim must be made within four years of the commencement of the LTA. In line with s 25(3) of the Constitution, according to s 23 of the LTA the owner of affected land is entitled to ‘just and equitable compensation as prescribed by the Constitution for the acquisition by the applicant of land or a right in land’.

The LTA, as underpinned by s 25, thus provides a substantively transformative framework for the redistribution of land. Unfortunately, there have been numerous problems with the administration of this component of the LTA, with few s 16 claims having been resolved (Msiza is a notable, but long-awaited exception). The government’s failure to settle some

86 HLPR (note 37 above) at 205.
87 This, according to s 1(xi)(c) of the LTA, includes a successor of a labour tenant but excludes a farm worker (farm workers are covered by the Extension of Security of Tenure Act 62 of 1997, discussed below). Juanita Pienaar and Jason Brickhill note that, although the purpose of the LTA is to phase out labour tenancy whereas the purpose of the Extension of Security of Tenure Act 62 of 1997 is to provide ongoing protection against the exploitation of farm workers, the continued distinction between labour tenants and farm workers is not ideal (J Pienaar & J Brickhill ‘Chapter 48: Land’ in S Woolman & M Bishop (eds) Constitutional Law in South Africa (2nd Ed, 2014)(CLOSA) 48-16.
88 LTA, s 1(xi).
10 914 labour tenant claims gave rise to protracted litigation brought by the Legal Resources Centre on behalf of the Association for Rural Advancement and various labour tenants to have a Special Master of Labour Tenants appointed to ensure that the Minister of Rural Development and Land Reform implements multiple court orders for redistribution of land under the LTA. On 8 December 2016, the LCC (which has primary jurisdiction over the LTA),\(^9^9\) ruled that, by refusing to implement labour tenants’ land ownership claims since the early-2000s, the Minister and Director General of the Department of Rural Development and Land Reform had acted inconsistently with the Constitution. It granted an order appointing a Special Master of Labour Tenants, who was tasked to produce a plan to implement all the amassed labour tenants’ claims to facilitate ownership in collaboration with the Department of Rural Development and Land Reform.\(^9^0\)

In its amended form,\(^9^1\) the PLAA is the main legislative mechanism to realise the land ownership redistribution programme. From a legal perspective, the PLAA provides an extraordinarily transformative framework in terms of which land\(^9^2\) is designated by the Minister of Rural Development and Land Reform for the purposes of redistribution to persons who have no land or restricted access to land, persons wishing to upgrade their land tenure or persons who have been dispossessed of their right in land but do not have a right to restitution under the Restitution of Land Rights Act 22 of 1994.\(^9^3\) As with the LTA, the PLAA has been under-utilised. More generally, the HLPR paints a bleak picture of government efforts to comply with s 25(5) of the Constitution.\(^9^4\)

Regarding land occupation and use, in recognition of the large numbers of black people who live (and often work) on land that someone else owns, s 25(6) establishes the need for the state to recognise and protect land occupation and use rights even where these rights clash with land ownership rights. It also requires Parliament to enact legislation to give effect to s 25(6), and several laws have been promulgated to this end. Chief among these legal reforms are the LTA, which (apart from enabling land acquisition) provides enhanced protection against the exploitation and eviction of labour tenants (ch II, LTA);\(^9^5\) the Interim Protection

---

\(^9^9\) The LCC also has primary jurisdiction over the Extension of Security of Tenure Act 62 of 1997 and the Restitution of Land Rights Act 22 of 1994, discussed below.

\(^9^0\) *Mwelase & Others v Director-General for the Department of Rural Development and Land Reform &Others* [2016] ZALCC 23, 2017 (4) SA 422 (LCC) (‘*Mwelase*’). Disappointingly, on 17 August 2018, the SCA overturned the appointment of the Special Master in *Mwelase* (on the grounds of judicial over-reach). However, the Legal Resources Centre has indicated it will seek leave to appeal the judgment to the Constitutional Court, available at http://lrc.org.za/news/press-release-disappointing-sca-outcome-labour-tenants-special-master-rejected/.

\(^9^1\) The PLAA, which was originally called the Provision of Certain Land for Settlement Act, has undergone various amendments and name changes over the years including: the Land Affairs General Amendment Act 11 of 2000, the Provision of Land and Assistance Amendment Act 58 of 2008 and the rural Development and Land Reform General Amendment Act 4 of 2011. Pienaar notes that since 2011, the PLAA has been known as the Land Reform: Provision of Land and Assistance Act 126 of 1993 (Pienaar (note 54 above) 287).

\(^9^2\) Both state and privately-owned land can be designated but privately-owned land can only be designated for this purpose if it has been made available by the owner (who would then receive compensation).

\(^9^3\) Pienaar & Brickhill (note 87 above) at 48–21.

\(^9^4\) HLPR (note 37 above) 205–231.

\(^9^5\) LTA, s 3(1) (stipulates that a person who was a labour tenant on 2 June 1995 shall have the right ‘with his or her family members to occupy and use that part of the farm’ that he or she was using and occupying on that date’) LTA, s 9(1) (establishes that a labour tenant who is over 65 years of age, or as a result of disability is unable personally to provide labour to the owner or lessee; and has not nominated a person to provide labour in his or her stead, ‘shall not be evicted’ for reasons of failure to work for the owner or lessee).
of Informal Land Rights Act 31 of 1996 (‘IPILRA’), which governs the tenure security of people who occupy and use land under customary land law;\(^96\) the Extension of Security of Tenure Act 62 of 1997 (‘ESTA’),\(^97\) which provides enhanced protection against the eviction of people who occupy non-formally proclaimed township areas with the consent of the owner;\(^98\) and the Prevention of Illegal Eviction from and Unlawful Occupation of Land Act 19 of 1998 (‘PIE’), which protects unlawful occupiers anywhere in South Africa from being evicted un-procedurally or unjustly.

As outlined in part II above, there are considerable problems with the regime governing customary land rights, including IPILRA. However, whatever the complexities of IPILRA and the broader frameworks for governing customary land law (including the controversial Traditional Leadership and Governance Framework Act 41 of 2003 and the highly contested TCB), it is not s 25 that blocks transformative change (however this may be defined) in the customary land rights arena. Regarding IPILRA itself, the HLPR laments the fact that ‘despite the constitutional imperative to secure and protect land tenure’, Parliament has failed to enact ‘substantive legislation to defend communal land tenure’, leaving the 17 million South Africans who live in the former homelands with only the interim ‘holding measure’ or ‘safety net’ of IPILRA.\(^99\) The HLPR recommends inter alia that IPILRA be made permanent, and that it is clarified that ‘holders of informal rights’ rather than the state (including traditional authorities) are deemed to be the owners of the land’ for the purposes of decision-making and revenue.\(^100\)

The courts have arguably done what they can in the absence of comprehensive legislation to defend customary communities against the erosion of their land rights by political attempts to bolster the position of traditional leadership.\(^101\)

The courts have also played a critical role in interpreting the scope of PIE and ESTA. Regarding ESTA, beyond affirming the additional protections against unjust evictions,\(^102\) the courts have established that ESTA protects a range of rights including the right of a wife to remain on the land notwithstanding the valid dismissal and eviction of her husband.\(^103\) The Court has also recently, in Daniels,\(^104\) ruled that where landowners refuse to improve their

---

\(^96\) IPILRA protects several categories of rights: use or occupation of, or access to land in terms of (i) any tribal, customary or indigenous law or practice of a tribe; or (ii) the custom, usage or administrative practice in a particular area or community where the land at any time vested in the South African Development Trust, the government of any self-governing territory or the former governments of the four national states.

\(^97\) In addition to security of tenure provisions, ESTA, like the LTA, provides for redistribution (ch II, ESTA). However, this form of land redistribution has hardly been invoked and where it has, it has not been very successful (Pienaar & Brickhill (note 87 above) at 48-16).

\(^98\) This includes a right to family life, not present in the Constitution (s 6(1)(d)), burial rights (s 6(4)) and safeguards against unfair eviction (ss 8–13). An ESTA Amendment Bill B224-2016 has been before Parliament for several years. The HLPR notes various problems with this Bill that, if approved, would result in an erosion of farm dweller’s rights (HLPR (note37 above) at 295–297).

\(^99\) HLPR (ibid) at 258.

\(^100\) Ibid at 270.

\(^101\) During 2018, the courts further reinforced the informal land rights of customary communities under IPILRA (over the rights of traditional authorities, the state and mining companies), in the cases of Malelu & Others v Itereleng Bakgatlha Mineral Resources (Pty) Limited & Another 2018 [ZACC] 41, 2019 (2) SA 1 (CC) and Baleni & Others v Minister of Mineral Resources & Others [2018] ZAGPHC 829, 2019 (2) SA 453 (GP). The latter case was under appeal at the time of writing.

\(^102\) Pienaar & Brickhill (note 87 above) at 48–37 and 48–42.

\(^103\) Klaase & Another v Van der Merwe NO & Others 2016 [ZACC] 17, 2016 (6) SA 131 (CC).

\(^104\) Daniels v Scribante & Another 2017 [ZACC] 13, 2017 (4) SA 341 (CC)(‘Daniels’).
accommodation, in terms of ESTA, ESTA occupiers have the right to make such improvements that are necessary to live in acceptable conditions. Following an amendment to ESTA in 2001 that provided for burial rights in certain circumstances, the LCC found in Nhlabathi\(^{105}\) that – contrary to the arguments of the landowner that the ‘appropriation of a grave deprives the land-owner of property’ – the right to burial was introduced by legislation as part of the constitutional mandate to ensure legally secure tenure,\(^{107}\) and that the appropriation of a grave in terms of ESTA amounts to a ‘minor intrusion’ only and not requiring compensation.\(^{108}\)

The right of family members to receive ESTA protection, the right of farm dwellers to live in adequate standards of accommodation, and the right to bury on the land regardless of the owner’s wishes, constitute significant and transformative inroads of use and occupation rights vis-à-vis ownership rights. In addition, in recognition of the typical vulnerability and low-income status of occupiers covered by ESTA, in Nkusi,\(^{109}\) the LCC interpreted ESTA as providing a right to legal representation at state expense for occupiers whose tenure is threatened or infringed.\(^{110}\)

Regarding PIE, in a series of cases, the courts have established a set of transformative principles that clarify not only the state’s obligations, but also (as flagged above in the analysis of deprivation of private ownership rights in the context of unlawful occupation) the constitutional and public interest-related duties of property owners and the limits of property ownership rights, in relation to unlawful occupiers:

- The state has an obligation to devise and implement a reasonable housing policy that, at the minimum, provides emergency shelter for those who would otherwise be rendered homeless by an eviction;\(^{111}\)
- Where an eviction would render occupiers homeless, it would ordinarily not be just and equitable to evict them without the state providing alternative accommodation;\(^{112}\)
- The state has an obligation to meaningfully engage occupiers regarding evictions;\(^{113}\)
- The state has the same obligation to provide emergency shelter to evictees who would otherwise be rendered homeless regardless of whether the eviction is initiated by the state or a private land owner;\(^{114}\)
- Where the provision of emergency shelter (whether directly or via a service provider) entails draconian rules such as gender-segregation and lockout during the day, this is an unjustified limitation of occupiers’ rights, and therefore unconstitutional;\(^{115}\)

---

\(^{105}\) Section 6(2)(dA) of ESTA now provides that an occupier ‘has the right to bury a family member who resided on the land at the time of his or her death, in accordance with their religion or cultural belief, if an established practice in respect of the land exists’. And, according to section 6(5), ‘family members of a long-term occupier have the right to bury the occupier on the land on which he or she was resident at the time of death.’

\(^{106}\) Nhlabathi v Fick [2003] ZALCC 9 (‘Nhlabathi’).

\(^{107}\) Ibid para 26(d).


\(^{109}\) Nkusi Development Association v Government of the Republic of South Africa 2002 (2) SA 733 (LCC) (‘Nkusi’).

\(^{110}\) Pienaar & Brickhill (note 87 above) at 48–36.


\(^{112}\) Port Elizabeth Municipality (note 55 above)

\(^{113}\) Occupiers of 51 Olivia Road, Berea Township and 197 Main Street, Johannesburg v City of Johannesburg & Others [2008] ZACC 1, 2008 (3) 208 (CC).

\(^{114}\) Blue Moonlight (note 65 above).

\(^{115}\) Dladla & Another v City of Johannesburg & Others [2017] ZACC 42, 2018 (2) SA 327 (CC).
• Where, due to the scale of the unlawful occupation, it is unfeasible to order an eviction because it would not be just and equitable to do so, the state might be compelled to purchase (or expropriate) the property from the property owner to fulfil its constitutional obligations;\textsuperscript{116}
• The state cannot use disaster management legislation as a way to evict occupiers without complying with PIE;\textsuperscript{117}
• Private owners seeking to evict unlawful occupiers bear the onus of showing that it is just and equitable to evict them (rather than the unlawful occupiers having to prove that it would not be just and equitable to evict them);\textsuperscript{118} and
• A private land owner might have to tolerate unlawful occupation for a considerable time until the state can provide emergency shelter.\textsuperscript{119}

Notwithstanding these progressive interpretations by the judiciary, while there is nothing in the law that explicitly determines it, there has been an inherently conservative tendency by the judiciary when it comes to balancing ownership with occupation rights ultimately to grant eviction orders once the state has provided alternative emergency shelter, even if this is blatantly unsuitable and unacceptable. However, although in most cases the courts have granted eviction orders following an order ensuring that emergency shelter (however dire) is provided, in a recent case the court found that it was not just and equitable to evict a group of people – including two pensioners, who had lived on the property for 44 years working for the previous owner – from a property purchased by the current property owner, who wanted to redevelop the property for higher-income persons.\textsuperscript{120} In addition, the High Court judgment in Fischer indicates a new frontier of transformative adjudication in cases in which eviction of unlawful occupiers from private property is not equitable.\textsuperscript{121} In this case, the High Court ordered the government – in circumstances where it would be patently unjust and inequitable to evict a large number of unlawful occupiers from private land – to enter into good faith negotiations with Mrs Fischer (the landowner) to purchase her property.\textsuperscript{122}

\textsuperscript{116} Fischer v Persons listed on Annexure X to the Notice of Motion and those persons whose identity are unknown to the Applicant and who are unlawfully occupying or attempting to occupy Erf 150 (remaining extent) Phillipi, Cape Division, Province of the Western Cape & Others; Stock & Others v Persons unlawfully occupying Erven 145, 152, 156, 418, 3107, Phillipi & Portion 0 Farm 597, Cape Rd & Others; Copper Moon Trading 203 (Pty) Ltd v Persons whose identities are to the Applicant unknown and who are unlawfully occupy remainder Erf 149, Phillipi, Cape Town & Others [2017] ZAWCHC 99, 2018 (2) SA 228 (WCC) (‘Fischer’). For an analysis of the Fischer judgment, J Dugard ‘Modderklip Revisited: Can Courts Compel the State to Expropriate Property where the Eviction of Unlawful Occupiers is not Just and Equitable?’ (2018) 21(1) Potchefstroom Electronic Law Journal 1–28.

\textsuperscript{117} Pheko v Ekurhuleni Metropolitan Municipality [2011] ZACC 34, 2012 (2) SA 598 (CC); Schnurb Park Residents’ Association v City of Tshwane Metropolitan Municipality [2012] ZACC 26, 2013 (1) SA 323 (CC).


\textsuperscript{119} President of the Republic of South Africa & Another v Modderklip Boerdery (Pty) Ltd [2005] ZACC 5, 2005 (5) SA 3 (CC); Blue Moonlight (note 65 above).

\textsuperscript{120} All Builders and Cleaning Services CC v Matlaila & Others [2015] ZAGPJHC 2.

\textsuperscript{121} The Fischer judgment has been appealed to a full bench of the Western Cape High Court, with no set down date at the time of writing this article. It should be noted that, at least from the perspective of race, the Fischer case presents a challenge. Mrs Fischer is an elderly black woman who is not well off. She and her family have lived on their property since 1969. This land is now occupied as part of a mass-occupation by black communities who could find nowhere else to settle. If the government expropriates Mrs Fischer’s land to fulfil its constitutional housing obligations towards the unlawful occupiers, this case will highlight the importance of having a nuanced framework for calculating compensation provided for in s 25(3) of the Constitution.

\textsuperscript{122} Fischer (note 116 above) para 197.
D  Section 25(7): restitution

Land restitution is probably the most complex and emotive of the land reform processes delineated in s 25. It has also been one of the least successful processes in terms of both the relatively low number of instances in which land has been restored to claimants and the questionable success of restitution where this has occurred.\(^{123}\) The HLPR emphasises that the evident failures of the restitution mandate relate overwhelmingly to multiple problems with the Land Claims Commission,\(^{124}\) which has not yet finalised the claims that were lodged by the cut-off date of 31 December 1998,\(^{125}\) and/or government implementation processes,\(^{126}\) rather than with the text of s 25(7) per se. These problems have resulted in the majority of successful claimants opting for relatively paltry cash settlements rather than pursuing restitution per se,\(^{127}\) a fact that has undermined the potential for the restitution process to contribute towards greater land redistribution.

In terms of the legal frameworks governing restitution, the scope and detail of s 25(7) have been fleshed out in the Restitution of Land Rights Act 22 of 1994 (‘RLRA’). The RLRA echoes the main features and requirements of s 25(7) but set the deadline for making any land claims at 31 December 1998.\(^{128}\) Although there have been some conservative judgments particularly from the LCC in the early years,\(^{129}\) the general arc of judicial interpretation of the RLRA has been transformative, and the courts have interpreted key concepts progressively. For example, in an early case, Kranspoort, the LCC rejected a narrow interpretation (which would have required a high degree of association and cohesion between often far-flung members of the erstwhile community) of the requirement (in both s 25(7) and the RLRA) concerning a ‘community’ dispossessed of property after 1913, arguing that it ‘would be a grave injustice if the RLRA is to be interpreted so that the tragic consequences of a removal become the main reason why a community restitution claim aimed at remedying the removal should fail’.\(^{130}\)

Another transformative clarification was provided by the Court in Popela, when it generously interpreted the words (again reflected in both s 25(7) and the RLRA) ‘as a result of past racially discriminatory laws or practices’ to mean ‘as a consequence of’ and not necessarily ‘solely as a

---

\(^{123}\) See HLPR (note 37 above) at 232–256. See also Cousins & Hall (note 10 above) at 1–2.

\(^{124}\) Systemic problems, as detailed in the HLPR, include, at the level of the Land Claims Commission: high staff turnover and insufficient training, ineffective bureaucratic processes (with files in many places in disarray) and corruption; and, regarding the claims themselves, conflicting and overlapping claims, often resulting in the ‘bunching’ of unrelated claimants together (HLPR (note 337 above) at 232–256).

\(^{125}\) According to the HLPR, there are still 19 000 non-finalised and 7 000 unsettled claims arising from the claims process that closed at the end of December 1998; at an average rate of 560 claims being settled per year, it will take approximately 35 years to process these claims (HLPR (note 37 above) at 233).

\(^{126}\) HLPR (ibid) at 232–256.

\(^{127}\) According to the Department of Rural Development and Land Reform in August 2017, of the approximately 80 000 claims settled to date, all but 7 478 claimants opted for cash settlements rather than land transfers: https://www.notesfromthehouse.co.za/opinion/item/54-questions-that-leave-more-questions-than-answers. See also B Atuahene We Want What’s Ours: Learning from South Africa’s Land Restitution Program (2014) 89–106

\(^{128}\) As outlined in part II above, President Zuma’s government had sought to re-open the restitution process via the RLRAA, but this was derailed by the Land Access Movement litigation.


\(^{130}\) In Re Kranspoort Community 2000 (2) SA 124 (LCC) (‘Kranspoort’) para 48. The HLPR recommends that if the RLRA is amended, it should adopt the definition of community from Kranspoort (HLPR (note 37 above) at 248).
This interpretation enabled more claims to satisfy the requirement, even where the dispossession/removal was not the sole consequence of past racially discriminatory laws or practices e.g. where other factors might also have played a role.

On the more contentious side, there is the thorny issue of the large number of cases in which restitution is deemed not feasible – mainly in urban areas and/or where there has been substantial subsequent development of the property. In such cases, the Land Claims Commission and courts have typically provided monetary compensation to claimants in lieu of restitution. The question arises as to whether the compensation in such cases can be regarded as ‘equitable redress’ as required by s 25(7) and the RLRA. A recent Constitutional Court case that illustrates the difficulties of awarding compensation in such situations is Florence.132 In this case – in which restitution of a piece of land in Cape Town was deemed not feasible on the ground that it had, in the more than 25 years since dispossession, been developed into a parking lot and shopping complex – the Court grappled with how to calculate compensation for the claimants. The Court settled on a formula which, summarised, is as follows:

- Take as the starting point the value of the property at the time of dispossession – in this case it had been valued at R32 000 in 1970. Subtract any monies received for the property – in this case R1 350 was paid to Mr Florence, leaving a shortfall of R30 750.
- Then, in order to capture ‘the changes in value over time’ as per section 33(eC) of the RLRA, add CPI (inflation), in this case establishing an amount of compensation owed of R1 488 890 (plus a court-awarded solatium of R10 000).133

At first blush, the amount of close to R1 500 000 seems like an attractive sum. However, it is less attractive when considered in light of the fact that the property had a current market value of almost ten times this amount. In defence of the Florence formula (which had been applied in a string of earlier cases in other courts and was applied by the LCC and SCA in the Florence litigation), it might be questioned whether, had they not been dispossessed, the claimants would have developed the property to achieve the same current market value. But what if they would have? And, regardless, can this amount of compensation be regarded as ‘equitable redress’ when compared with the current market value of the property? Or does this constitute what Ms Florence argued was ‘an illogical discrepancy between the value of restitution in the form of restoration and restitution in the form of financial compensation’?134 The Florence formula for calculating compensation in instances where restitution is not feasible raises questions about the adjudication of restitution claims, especially in light of the de facto awarding of current market value to current owners in

133 In a dissenting opinion (Cameron J, Froneman J and Majiet, AJ concurring), Van der Westhuizen J would have awarded R2 211 732.54, which reflects the amount had the under-compensated amount been invested in a 32–day notice deposit facility since the day of the LCC initial judgment in the matter. A separate dissenting opinion, by Froneman J, comes closest to acknowledging the underlying injustice of the discrepancy in compensation: ‘I find it difficult, however, to conceive how one can ever adequately determine proper compensation for people who were forced to sell their property at the time of dispossession. The only way to compensate them for their loss is restoration or, if restoration is not possible, something as close as possible in financial terms to restoration. In this kind of case the CPI, as representing the change in the value of money, may be inappropriate, and the present market value of the property could serve as evidence of its inappropriateness’ (Florence (ibid) para 203).
134 Florence (ibid) para 147.
land reform contexts. Responding to this apparent puzzle, the HLPR recommends that ‘the question of equitable redress needs to be revisited’.135

In general, the courts and especially the Constitutional Court have pursued substantively transformative interpretations of the legal frameworks governing restitution. A notable exception is the Court’s endorsement of the asymmetrical and unfair discrepancy between the de facto current market value regime (as pursued at least until 2018) for compensating current owners and the restricted formula for compensating historically dispossessed persons in instances where restoration is deemed not feasible. However, as an incident of judicial interpretation, this approach could be shifted through new legislation or subsequent Constitutional Court interpretation.

E Section 25(8): any other measures

As highlighted in this article, it would be possible, legally, for the government to pursue a much more substantively transformative land reform programme than it has pursued until now. But, returning to the question left hanging in the analysis of s 25(2), (3) and (4) above, would it be possible for the government to bypass the compensation clause entirely? The requirement for ‘just and equitable compensation’ according to s 25(3)’s matrix, arguably provides an appropriate and substantively transformative method for balancing the various rights and considerations and achieving what, in some circumstances, might be nil compensation. However, if this is not considered to be sufficiently transformative a premise for government action, tantalisingly s 25(8) of the Constitution, which has not received much attention and has yet to be conclusively litigated, states: ‘No provision of this section may impede the state from taking legislative and other measures to achieve land, water and related reform, in order to redress the results of past racial discrimination, provided that any departure from the provisions of this section is in accordance with the provisions of s 36(1)’. This suggests it is likely that, should action be pursued, and/or a law be adopted that enables the state to expropriate property for the purpose of land restitution without any compensation, this could be deemed constitutional if found ‘reasonable and justifiable in an open and democratic society based on human dignity, equality and freedom’. Thus, s 25(8) signals that, notwithstanding any of the limits discussed here that relate to the constitutional or legislative frameworks or judicial interpretation thereof, it would be possible for any government to pursue quite radical land reform and for this to fall within the legal parameters of s 25.

IV CONCLUSION

Sidestepping scholarly and popular debates about whether s 25 of the Constitution is optimal or should be amended, this article has focused on what can be achieved within s 25’s existing formulation. Specifically, the article has empirically examined the legal limits of any potential project of transformative land reform, as understood within a substantive (rather than revolutionary) model of equality. The article has argued that, while recognising different regimes and kinds of property rights including private ownership rights, s 25 provides a permissive – and even mandatory – template to pursue transformative land reform. Certainly (in the same way as the Domestic Violence Act 116 of 1998 does not eradicate domestic violence), s 25

135 HLPR (note 37 above) at 250.
does not automatically ‘undo the settler-created house’. But it allows expropriation in the public interest and for a public purpose that explicitly includes redistributive reforms. It also establishes a mechanism for individuals and communities to secure the return of historically dispossessed land. And it sets out a matrix for calculating compensation to current owners that, on a proper application, could amount to zero compensation in appropriate cases. As such, s 25 provides a framework within which much more could be done to alter land ownership, occupation and use patterns. Indeed, it is hard to visualise how (even in a revolutionary equality model) in practical terms the ‘settler-created house’ could more justly be deconstructed given the overlapping and conflicting land rights claims etched in the land over centuries of use, occupation, ownership and dispossession.

For their part, the courts have played an important role in interpreting s 25 in transformative ways. Nonetheless, the courts could certainly do more to push the boundaries of substantive transformation. In particular, the courts could play a more robust role in deliberations over compensation – by unravelling the de facto current market value paradigm pursued by the government until now, and by pursuing a more equitable formulation for compensation for historically dispossessed individuals or groups when restitution is not feasible. In addition, courts could develop a stronger approach to hierarchical clashes of rights especially as between owners and unlawful occupiers. Recent cases, including Fischer suggest that the courts are moving in this direction. However, there is ultimately a very real limit to what the courts can achieve in the absence of political direction. As pointed out by Geoff Budlender in 1992,

The only way to achieve a true balance between the rights of property-holders and property-less is to weaken existing property rights, as a matter of deliberate policy. Whether that should and could be done is a question not of constitutional principle, but of political power and priorities.

Finally, in relation to the potential for land restitution specifically – and possibly land reform more generally – to resolve South Africa’s daunting problems of historical injustice and socio-economic inequality, it is sobering to read the cautionary note from Andries du Toit as quoted in the Kranspoort judgment:

This brings us to the most pressing and painful part of the problem – which is that the moment of return to the land cannot live up to the expectations and hopes generated by it. For of course what was lost can never be returned. Part of the problem is that the land is not the only thing that was lost. What was destroyed through … removals was a whole way of being, a set of community relations, a system of authority and let [us] not forget, a broader system of economic relations and livelihoods of which the land was but a part, and which gave it its function and its value. The terrible truth of Restitution has thus been that the moment of return to the land is often a moment of disappointment and anti-climax. To settle on the spot from which one’s forebears – or even a younger, more vigorous, more hopeful self – were once removed, is not necessarily to return to that more authentic, more dignified, more hopeful mode of existence. As we have seen in numerous cases … to return to face the complex, dispiriting and painful problems of the new South Africa once again in new and often more intractable ways. For communities have grown, services are needed and the rural and national economies that made certain forms of existence possible may no longer be in palace. If existence without piped water

---

and electricity was acceptable in the past, it is no longer so – and these services have to be paid for, and paid for in a very different, increasingly globalised, economy. In all too many cases, we may be looking at a scenario where the land is returned to those who lost it – only to be lost again to the banks, or to those who are willing to pay good cash for it.\textsuperscript{138}

From Housing to City: On the Possibilities of the Right to the City in South Africa and India

AJEY SANGAI

ABSTRACT: India and South Africa face a housing crisis where millions of people are either homeless or live in grossly inadequate conditions exposed to the elements of nature and severe health and nutrition risks. The crisis affects people from vulnerable races, castes, religious communities and other disadvantaged groups more severely than others. This paper argues that (i) this crisis has a spatial context which is manifested in the production of grossly inegalitarian cities that has pushed these disadvantaged groups to the margins, and (ii) homelessness and underserved housing is not natural but a consequence of a history of state actions through laws, regulations and even judicial orders. The paper focuses on the notion of the ‘right to the city’. Initially, an attempt is made to provide some content to this right which, I argue, could be understood in terms of (i) access to amenities in a city like schools, hospitals, market etc.; (ii) the right to be a participant in the decisions of the city; and (iii) the right to appropriate the opportunities and advantages of a city. While neither South Africa nor India recognises this right explicitly, both have a fertile constitutional jurisprudence on the right to housing. The paper takes housing jurisprudence as its point of departure and analyses the judgments of the Constitutional Court of South Africa and the courts in India from a ‘right to the city’ perspective. The analysis requires placing housing within a broader sociological and historical context and touches on the aspects of access, participation and appropriation. I argue that we should focus not only meeting the welfare-needs of disadvantaged groups but also on their wider citizenship-based claims to play a part in the cities in which they live which, importantly, also provides a response to historical injustices in both societies.

KEYWORDS: right to housing, right to the city, participation, corrective justice, India, South Africa

AUTHOR: Assistant Professor, Jindal Global Law School, O.P. Jindal Global University, Sonepat, India. Email: asangai@jgu.edu

ACKNOWLEDGEMENTS: I would like to thank the participants of the Constitutional Court Review IX Conference for their feedback. I received generous comments and questions from the peer-reviewers that helped me organise my paper better. Specifically, I am grateful for the encouragement and support of David Bilchitz and Mispa Roux. I received vital suggestions from Jason Brickhill and Stuart Woolman. Ralph Madlalate sent me important essays on the topic published in South African journals. I also thank
Swagat Baruah, a student of Gujarat National Law School, India, for research assistance. I would also thank my employer, O.P. Jindal Global University, Sonipat, Haryana, India for their support which helped me attend the conference. Apologies, if I missed anyone.
I INTRODUCTION

‘It is harder in 2013 to reverse the apartheid geographies than it was in 1994’, claims the State of South African Cities Report, pointing to a crisis that is as fundamental as it is comprehensive. Further, the report notes that unlike the Global North, urbanisation in African cities is characterised not by industrialisation, but by the ‘urbanisation of poverty’ where the failure of the cities to provide a structural transformation to match the economic and demographic changes has left millions of denizens and migrants vulnerable and powerless.

The United Nations had estimated that over 100 million people have no place to live and over one billion people are living in grossly inadequate conditions. In addition to this, the alarming increase in global eviction rates around the financial crisis and other risks that evicted and homeless people are exposed to, we are looking at a crisis that is not only restricted to housing but is a product of grossly inegalitarian cities. This is exemplified by spatial inequalities visible in the differentiated patterns of housing. The poor and vulnerable are pushed to the underserved margins of the cities which makes access to resources unequal. While the growth of suburbs shifted the centre of the city itself, it left the existing dwellers who had settled and built their lives around serving the needs of previous city centres, helpless and disenfranchised as economic opportunities shifted to new centres. As will be explained, it would be incorrect to regard spatial inequalities as natural and inevitable, but an impoverishment actively brought upon by the policies of the state, and often aided by the judiciary. The crisis of housing, eviction and locational inequalities, along with the laws that criminalise the homeless, reveal an important anomaly – denial of private shelter and exclusion from public space – that compels a deeper look into the processes that constitute a ‘city’

2 Ibid 24. The urbanisation of poverty is characterised by the lack of structural transformation of cities and consequent denials of a number of inter-connected human rights to the urban poor who characteristically reside in the margins of the town in underserviced housing. G Piel ‘The Urbanization of Poverty Worldwide’ (1997) 40 Challenge 58. Also, M Ravallion, S Chen, & P Sangraula ‘New Evidence on Urbanisation of Global Poverty’ (2007) 33 Population and Development Review 667: ‘The negative externalities of geographically concentrated poverty and irreversibilities resulting from the costs of migration, which can mean that migrants to urban areas cannot easily return to their former standard of living in rural areas’ means that urbanization, rather than being a harbinger of growth, is an ‘unwelcome forebear of new sources of poverty.’
4 S Soederberg ‘Eviction: A Global Capitalist Phenomena’ (2018) 49 Development & Change 286. In addition to this, studies have explored the link between eviction rates and suicide, where an evictee is over four times more likely to commit suicide than others, even if other suicidogenic factors are controlled. Also, Y Rojas & S Stenberg ‘Eviction and Suicide: A Follow-up Study of Almost 22000 Swedish Households in Wake of Global Financial Crisis’ (2016) 70 Epidemiol Community Health 409.
5 G Bhan ‘This is no longer the city I once knew’: Evictions, the Urban Poor and the Right to the City in Millennial Delhi’ (2009) 21 Environment & Urbanization 127.
Despite being on the legal agenda at both global and national levels, the right to housing has only recently become subject to normative analysis in academic literature. While situating access to adequate housing in the landscape of the city has been a niche area of the urban studies and legal geography scholarship, constitutional lawyers and comparativists have also not had much to say about this. For nations with commonalities in their histories that enabled migration of constitutional ideas, there is little dedicated comparative study on the right to housing in India and South Africa, though there are individual contributions in adopting the right to the city framework to understand the adequacy of housing rights. Curiously, India has formally opposed the inclusion of the right to the city in the draft New Urban Agenda (that will define the way in which cities worldwide are shaped over the next two decades) even as India aspires to build 100 new ‘smart cities’. This is partly due to the imperfect understanding of the legal meaning and implication of this right which distinguishes itself from other rights in its people-centric approach and partly as a result of a genuine scepticism about the democratic potential of this right amidst the need for rapid economic growth. This is in stark contrast with Latin American nations, for example, that have given legal recognition to this right.

The right to the city includes issues relating to access, opportunity, participation, and capability to exercise ones membership in the city. However, given that it has been invoked frequently in the contexts of evictions, displacement, homelessness, locational disparities and spatial justice, housing rights become an integral component of this right. Of course, one need not reside in the precise jurisdictions of a city to access it. While this right is still somewhat amorphous, part II of this article will try to provide a working conception of this right. This part (I) has two objectives. First, it describes the housing crisis that faces both the nations. Second, it explains the role of the state in bringing about that crisis, and this connection makes the linkages between the housing rights and the right to the city more explicit. It debunks the idea that the existing arrangements of the Indian and South African cities are natural. However,

---

10 L Weinstein & X Ren ‘The Changing Right to the City: Urban Renewal and Housing Rights in Globalising Shanghai and Mumbai’ (2009) 8 City & Community 407 is a comparative study, yet it concentrates primarily on regulations and policy instead of legislations and judgments.
given that this right to the city has not really found judicial recognition in either India or South Africa, part III of the article begins with a detailed comparison of the right to housing in South Africa and India. The aim is to analyse the extent to which it is possible to find an implicit recognition of the right to the city through the interpretation of the right to housing in these jurisdictions. A regressive judicial order on the right to housing would also put fetters on the right to the city. Finally, this article concludes by taking stock of the developments that the courts of both the nations have made towards recognising a right to the city and the challenges that lie ahead for these two nations in their pursuit of creating smart and world class cities.

II CONCEPTUALISING THE RIGHT TO THE CITY

David Harvey, poignantly says, The Right to the City is far more than the individual liberty to access urban resources: it is a right to change ourselves by changing the city. It is, moreover, a common rather than an individual right since this transformation inevitably depends upon the exercise of collective power to reshape the process of urbanization. The freedom to make and remake our cities and ourselves is … one of the most precious yet most neglected of our human rights.

The statement suggests two possible conceptions of this right, incremental and transformative. The incremental approach suggests that the right to the city is essentially a bundle of rights to access various amenities and resources a city has to offer. The transformative approach, on the other hand, makes the right to the city an existential and citizenship-based claim entailing consequences for an urban democracy in the light of significant changes in the general governance of the cities. Succinctly put, the incremental approach defines it as a collection of rights in the city. The transformative approach takes this right to be larger than the sum of its parts.

For understanding the transformative approach, Lefebvre is helpful. He saw this right as a cry and a demand of the inhabitants for 'sharing in the fullness of the urban life.' He conceptualised the city as oeuvre or work in progress whose inhabitants not only make claims for inhabiting the city but also for appropriating the experience it has to offer and participating in the evolving forms of city life and cityscapes. The transformative approach thus entails two types of rights: (i) the right to participate (which includes exercising one’s franchise and

---

13 Interestingly, the High Court of Delhi recently engaged with the concept of the right to the city in a case pertaining to the eviction of 5 000 families of slum dwellers, the Jhuggi-Jhopri Colony of Delhi. Ajay Maken v Union of India, W.P. (C) 11616 of 2015, decided on March 18, 2019 (Ajay Maken').
14 D Harvey 'The Right to the City' in Social Justice and the City (Rev ed, 2009) 315.
15 Surveying the literature on the right to the city and urban governance, Mark Purcell outlines three major shifts in urban governance which has increased the disenfranchisement of city people and denied them opportunities to appropriate the amenities and experience a city offers. These shifts are: (i) rescaling of urban governance; (ii) policy reorientation in favour of competition over redistribution; and (iii) transfer of several state functions to non-state or quasi-state body in shift of attitude from government to governance. M Purcell 'Excavating Lefebvre: The Right to the City and its Urban Politics of the Inhabitants' (2002) 58 Geojournal 99, 100–101.
16 UNESCO – Right to City in the Indian Context, 2–3. The authors of this paper preferred to call the incremental approach as reformist and the transformative approach as radical. However, I believe that ‘incremental’ and ‘transformative’ better captures the core motivation informing the divide.
17 H Lefebvre 'Right to the City' in E Kofman & E Lebas Writings on Cities (trans 1996). Also, Coggin & Pieterse (note 11 above) 259–260.
18 Ibid.
19 Purcell (note 15 above) 102–103.
agency in important collective decisions of urban life, and not merely participation in the state’s decision-making); and (ii) the right to appropriate what the city has to offer (which not only includes the right of the inhabitants to physically access, occupy and use the urban space but also a right to ‘produce’ the urban space that meets the needs of the inhabitant). For Lefebvre, therefore, the use value of the city must trump the exchange value that commodifies it. A city viewed from the exchange value perspective appears as site for accumulation where private property rights trump other social concerns of existence. Contrarily, the use value perspective does not deny the importance of property rights but questions their primordiality over other concerns such as shelter, food and livelihood. It sees the city as a site of inhabitation instead. Prioritising use-value, thus, resists the preferential treatment of property rights of owners and emphasises the use-rights of the inhabitants. This radical conception may sometimes be disconcerting, for we are unsure what kind of city it would produce.20

A Concretising the right

The right to the city is a complex right. It combines legal and moral, formal and substantive, and individual and collective rights. Coggin and Pieterse argue that the right to the city is not reducible to a ‘legal notion of a right, being something fairly stable, clear and precise, that can be consistently and predictably invoked, interpreted and enforced.’21 The chaotic nature of this right makes it difficult for constitutionframers, legislators and policy-makers to lay out a definitive content of this right.22 This paragraph briefly surveys how this right has been enunciated in legal instruments like legislation, charters and judgments, so as to arrive at a working conception of it. The World Charter for the Right to the City23 provides a helpful elucidation of the normative content of this right. It defines the right to the city as ‘the equitable usufruct of cities within the principles of sustainability, democracy, equity, and social justice.’24 Further, the Charter states that this right is a collective right of all the inhabitants of the city, particularly the vulnerable and marginal groups. It enables the inhabitants to ‘achieve full exercise of the right to free self-determination and an adequate standard of living.’25 The Charter also states that the right to the city ‘is interdependent of all internationally recognised and integrally conceived human rights.’ The World Charter, echoing Lefebvre, highlights citizens’ right to participate and appropriate. Not only does it provide for progressive realisation of all human rights including the right to clean environment,26 it also stresses the democratic

20 Ibid 100, 103.
21 Coggin & Pieterse (note 11 above) 262.
22 There have been recent attempts to lay down a constitutionally or legislatively protected right to the city. For example, the recently adopted Constitution of Mexico City in art 12 defines right to city as ‘the full use and the full and equitable enjoyment the city, founded on principles of social justice, democracy, participation, equality, sustainability, respect for cultural diversity, nature and the environment.’ Drawing from Lefebvre, the 2001 City Statute of Brazil, in recognising the social function of the city prioritises the ‘use value of the city’ over the exchange value. It also emphasises the democratic city management which is defined as ‘a path to plan, produce, operate and govern cities subject to social control and participation.’ The commentary is available at http://www.citiesalliance.org/node/1947.
24 Ibid art I(2).
25 Ibid.
26 Ibid art I(6).
governance\textsuperscript{27} of the city and the priority of its social function. The social function prioritises the use-value of the city over its exchange value by guaranteeing citizens the full usufruct of its resources and subordinating individual property rights.\textsuperscript{28}

The European Charter on Human Rights in the City\textsuperscript{29} lists principles and rights involved in the governance of cities. It provides for six principles, namely, (i) equal rights and non-discrimination;\textsuperscript{30} (ii) effectiveness of public service;\textsuperscript{31} (iii) transparency;\textsuperscript{32} (iv) subsidiarity;\textsuperscript{33} (v) solidarity;\textsuperscript{34} and (vi) international municipal cooperation.\textsuperscript{35} Based on these principles, the European Charter lays down the rights of the citizens. The right to the city provides inhabitants with conditions that allow them a sense of fulfilment from a ‘social, political and ecological point of view, while assuming (their) solidarity duties’. It also lists, the ‘right to political participation’, where the residents not only have the right to represent, but also to access public spaces for meetings and gatherings and freedom of forming associations, expression and demonstration. It also provides that citizens should have the right to municipal social protection policies to be provided on a non-commercial basis.\textsuperscript{36} The European Charter also provides for the right to ‘harmonious and sustainable urban planning’, which entails citizens’ involvement for orderly development of cities that respects environment and heritage.

Both the European and the World Charter emphasise pluralism, non-discrimination and solidarity, and enjoin the state to enact suitable laws and policies towards realizing all the constituents of the right to the city. In fact, the 2001 City Statute of Brazil also emphasises the same principles. It provides instruments to the municipality for various forms of social intervention for free use of private property, tenure regularisation for informal properties, and stimulating urban development with redistribution of collective benefits of urbanisation. It secures the right to participate by mandating urban housing democratisation, legal aid and protection to vulnerable groups and creating avenues for referenda and plebiscites for popular decision-making.\textsuperscript{37} Like the Charters, the City Statute also emphasises environmentally sensitive urbanisation. Following this law, the Brazilian Government established a Ministry for Cities.

\begin{footnotesize}
\begin{itemize}
\item\textsuperscript{27} Ibid art II(1).
\item\textsuperscript{28} Ibid.
\item\textsuperscript{30} The European Charter obligates local authorities to ensure that the rights under the Charter are equally available to all the residents of the city, without discrimination (ibid 4).
\item\textsuperscript{31} The local authorities should ensure effective public service is adapted to people’s needs. It should assess its service and also prevent situations of discrimination (ibid 4).
\item\textsuperscript{32} Local authorities should publish their principles and the minutes of meetings in accessible manner so that people are aware of their rights and duties (ibid 4).
\item\textsuperscript{33} The aim of this principle, the European Charter says is ‘to ensure public services are accountable to the authority closest to ordinary citizens and, therefore, more effective.’ (Ibid 4).
\item\textsuperscript{34} The local authorities should work together with citizen organisation to foster solidarity networks among their inhabitants. (Ibid 5).
\item\textsuperscript{35} This principle helps in sharing of knowledge among different municipal corporations across globe as far as possible. (Ibid 6).
\item\textsuperscript{36} Other rights listed in the European Charter include, right to education, health, work, leisure, culture, decent housing, privacy and protection of family life (Ibid 6–12).
\item\textsuperscript{37} ‘The City Statute of Brazil: A Commentary’ 95 (‘Global Platform’), available at https://www.ifrc.org/docs/idrl/945EN.pdf.
\end{itemize}
\end{footnotesize}
The Global Platform for the Right to the City also provides a useful matrix to understand this right.\(^{38}\) It identifies ‘spatially just resource distribution, political agency (particularly for women and other marginalised groups) and socio-cultural diversity’ as important pillars of this right. It lists the following components of this right: ‘a city with inclusive economies, cultural diversity, quality public spaces, enhanced political participation, gender equality, inclusive citizenship, free of discrimination and sustainable.’\(^{39}\) The document calls it a collective and diffused right,\(^{40}\) like the right to environment and culture. This right not only includes intra-generational equity but also inter-generational equity. It views the city as commons, meaning ‘all the inhabitants should have the capacity to equally access the urban resources, services, goods and opportunities of city life; and participate in the making of the city.’\(^{41}\) This again reinforces and clarifies the transformational as well as access-based incremental aspects of the right to the city. It clarifies the transformational component by elucidating the right to participate in and appropriate the advantages of city.\(^{42}\)

It could be observed that the incremental (right to access urban amenities) and transformative approaches complement one another.\(^{43}\) The transformative agenda appeals to social and political movements from below while the incremental approach takes cognisance of institutional improvements required to enhance access to urban amenities. While the right to the city aggregates several existing human rights, it also adds an important dimension of spatiality\(^{44}\) as it insists on the need to implement these rights and principles in ‘cities and human settlements from an interdependent, interrelated and indivisible approach.’\(^{45}\)

The Delhi High Court in a recent case concerning the forced eviction of 5000 slum dwellers endorsed the ‘common good’ conception of the city proclaimed by the World Charter.\(^{46}\) It held that in the context of a right to housing and shelter it was important to note the growing recognition of the right to the city which in turn acknowledges the contribution of slum dwellers in the social and economic life of the city. The right to the city would regard it as important that the slum dwellers have the right to make use of the resources and opportunities of a city life. The right is violated when a healthy urban life of some citizens comes at the expense of indignities and deplorable conditions in which a large number of slum dwellers live. This reinforces the spatiality of injustices in the cityscapes.

---


\(^{39}\) Ibid 2.

\(^{40}\) Ibid 3.

\(^{41}\) Ibid.

\(^{42}\) Ibid 3, 9.

\(^{43}\) UNESCO – Right to City in the Indian Context, 3.

\(^{44}\) For explanation of the concepts of space and spatiality, see D Massey ‘Philosophy and Politics of Space and Spatiality: some consideration’ (1999) 87 Geographische Zeitschrift 1. Massey conceptualises ‘space’ as always a product of interrelations which is then predicated upon the existence of plurality: ‘space’ is constantly in the process of becoming through the material processes that are necessarily embedded in it (at 2). This enables us to change ourselves by changing the city, as Harvey says (Harvey note 14 above). The upshot of this is that, according to Massey at 3 ‘(there is) a parallel between the manner of conceptualising space and the manner of conceptualising entities/identities (such as political subjects) but also space is from the beginning integral to the constitution of those political subjectivities.’ Given the dimensions of interrelations, plurality and becoming, Massey argues that space has always had an element of unpredictability or chaos (at 8–9). This, perhaps also makes the right to the city, as Coggin & Pieterse (note 11 above) argue, chaotic.

\(^{45}\) Global Platform (note 38 above) 6–7.

\(^{46}\) Ajay Maken (note 13 above).
The above discussion suggests a certain consensus that could help us formulate the initial content of the right to the city. A reasonable recognition of the right to the city would guarantee the right to self-determination and to participate in the decisions of the city and ensure access to amenities and advantages of the city on a non-discriminatory basis, besides creating inclusive and sustainable cities. As it includes the existing human rights within the context of spatial justice, the right to the city exemplifies the indivisibility and inter-dependence of human rights. The constitutions that encapsulate such integratedness could enable a robust construction of this right. The constituents of the right to the city are exercised in the context of spatiality. The space and the diverse people that constitute it co-exist. The inhabitants not only relate with each other, they also define their relationship with the space they inhabit, which provides them finite resources and opportunities for their functioning. Hence, the idea of what it means to live in a certain city is a constantly evolving one, depending upon where the people stood with respect to each other and their city. These rights are therefore exercised both collectively and relationally. Further, the evolving nature of the urban space requires that the meaning of these rights require constant re-examination and in the light of the history and sociology of a space, these rights should also be interpreted purposively.

Nonetheless, one generally accesses the city through one’s house and/or en route to one’s work (broadly construed). In fact, in Ajay Maken, the court calls the right to the city an ‘extension of the right to housing’. Therefore, the following paragraph discusses the relationship between the city and the housing.

B  The right to the city and the right to housing

The genesis of the right to the city movement, the 1871 Paris Commune, emphasised residence as the basis for urban citizenship and was fundamentally moved by a crisis of habitation that raised issues of citizen participation, and a more equitable control of urban infrastructure, especially housing. Holston notes that urban citizenship as the ground for the right to the city has reappeared relatively recently, not in Europe, but rather with the ‘intense urbanisation of the Global South.’ One feature of this urbanisation is the persistent housing crisis.

The Special Rapporteur on the Right to Adequate Housing notes – ‘Housing is the basis of stability and security for an individual or family. The centre of our social, emotional and sometimes economic lives, a home should be a sanctuary; a place to live in peace, security and

---

47 Coggin & Pieterse (note 11 above) 262.
48 See, A Barak ‘Purposive Interpretation in Law’ (2005). Barak uses Gadamer’s work on Hermeneutics to explain purposive interpretation. He rejects the synchronic approach of legal formalism and emphasises the historical-evolutionary and contextual meaning of law. This makes the purposive approach more suitable to interpreting the right to the city. A sociological and historical understanding of the context preceding and prevailing when the law was enacted could serve to determine the purpose behind the enactment.
49 This ‘partitioned’ view of city as a defined and contained place is resisted by Massey (note 44 above at 8). This would mean that one’s right to the city begins not only upon entering the jurisdictions of the city but from before. This alludes to Lefebvre’s idea of city that is lived and experienced (see, Purcell note 15 above). This means that the locations of centre and periphery need to be read more critically.
50 J Holston ‘Housing Crises, Right to the City and Citizenship’ in E Murphy & N Hourani (eds) The Housing Question: Tensions, Continuities and Contingencies in the Modernist City (2013) 255, 259. Holston writes: ‘In the case of the city, the right to housing is one of the substantive aspects of urban citizenship, part of the bundle of rights that belong to urban citizens.’
dignity. The Special Rapporteur prioritises the use value of housing over the exchange value, by explaining it as a right over a commodity. It is emphasised by juxtaposing forced eviction and uprooting of people from their communities with making space for luxury living and corporate real estate investments. This has made even basic housing in cities unaffordable for millions of people. This article takes the housing right as the point of departure to discuss the more ambitious right to the city. Insofar as housing rights are concerned, drawing from the above analysis, the right to the city dimension includes questions of affordability in the context of finance deployed as a proportion of income and debt towards housing, public participation in urban governance including the questions of (re)settlement and displacement, barriers to access and mobility, non-discrimination in housing, and access to amenities and transport, among other things.

In terms of international law, art 11 of the ICESCR enjoins state parties to recognise the right to an adequate standard of living which includes food, clothing and housing and continuous improvements in living conditions. General Comment No. 4 expands the content of the right to housing to include ‘the right to live somewhere in secure, peace and dignity.’ Despite the inherent contextuality of ‘adequate housing’, the General Comment has established certain baseline factors such as legal security of tenure; availability of infrastructure and amenities; affordability; habitability; accessibility (particularly from the standpoint of disadvantaged groups); location and cultural adequacy. General Comment No 4 underlines the indivisibility and integrity of civil and political rights and socio-economic rights and mandates the state parties to ensure that certain provisions are immediately secured (particularly those which require the state to exercise a duty of avoidance and encourage self-help practices among people). While the significance of housing among other human rights, some of which may

---


52 Ibid. The Special Rapporteur notes, ‘In developing economies, often informal settlements or long existing neighbourhoods located in ‘prime land’ are subject to evictions and displacement to make way for speculative investment. Residents are often rendered homeless, replaced by luxury housing that often stands vacant.’

53 C Berger ‘Beyond Homelessness: An Entitlement to Housing’ (1991) 45 University Miami Law Review 315, 316–317. See also, CESCR General Comment No. 4 – The Right to Adequate Housing: Article 11(1) of the Convention, E/1992/23 (66th Session, 1991) (‘General Comment No. 4’). Even the general comment notes that ‘personal or household financial costs associated with housing should be at such level that attainment and satisfaction of other basic needs are not threatened or compromised.’ (Para 8) In the same vein, tenants should be protected from exorbitant rent levels.


55 See, WHO Health Principles of Housing (1989), available at http://apps.who.int/iris/handle/10665/39847. Housing is adequate when it promotes the physical and mental wellbeing of its inhabitants.

56 Housing should be located where the residents can access livelihood, health care, schools, child care and other social facilities.

57 Modern technologies of housing should be culturally sensitive and the provision of housing facilities should ensure that the expression of cultural identity is promoted and not sacrificed.

58 See, WHO Health Principles of Housing (1989), available at http://apps.who.int/iris/handle/10665/39847. Housing is adequate when it promotes the physical and mental wellbeing of its inhabitants.

59 General Comment No. 4 (note 53 above) 4–5.
pertain to the right to city, is acknowledged, the approach adopted in the General Comment No 4 remains state-centric, even as it briefly alludes to the ‘duty of avoidance’. The right to the city envisages horizontality in the enjoyment of these rights. Under the right to participate in and appropriate the ‘usufructs of the city’, as discussed in the previous paragraph, there is a clear recognition of individual and collective agency and responsibility. The state-centric approach does not appreciate the need for direct participation of people in the major decisions regarding their location and engagement with the city, including the right to housing. The stories of displacement and resettlement that arise during developmental projects, including the creation of new urban and business spaces, re-inforce scepticism about the approach that relies wholly on a state dispensation. The right to the city calls for understanding the relationship between the right to participate and the right to appropriate as well as the incremental and transformational approaches, discussed above. In a sense, the right to the city is, as Baxi notes, a ‘right to struggle for maintaining critical social solidarities.’

The paragraphs above have laid the groundwork for the content of the right to the city through its philosophy and legal recognition, and provided justifications for starting with the right to housing. The following paragraphs contextualise the above discussion. I start with

---

61 Horizontality changes the conventional understanding where human rights laws placed duties on the state to protect and fulfil the human rights of individuals. Under this understanding human rights are vertically aligned. Under the horizontal conception, the private members as well as corporations have obligations to protect and fulfil the human rights of fellow members. See, JH Knox ‘Horizontal Human Rights Law’ (2008) 102 American Journal of International Law 1.

62 The ‘usufructs of the city’ would mean the resources and other advantages made available by the city for the use and benefit of its inhabitants. They could include amenities, a city life, social and economic opportunities, neighbourhood, access to schools and hospitals, etc.

63 The idea of agency and responsibility in the right to the city, as expounded in Lefebvre (note 17 above), Harvey (note 14 above), Purcell (note 15 above) and Massey (note 44 above) can be understood as coming from the existentialist philosophical tradition. For example, see JP Sartre Existentialism is Humanism (1977).

64 One example of perils of state-centric approach is noted by Chris de Wet in the context of development-based displacement, ‘when it comes to upholding the rights of resettled people vis-à-vis their government, the state is both player and referee’, C de Wet ‘Economic Development and Population Displacement: Can Everybody Win?’ (2001) 36 Economic & Political Weekly 4637, 4640–4641. The need for democratic participation is stressed in failed resettlement projects: ‘The result is that resettlement, which should involve the planned movement of people in such a way that provision is made for sustainable livelihoods in the new area, tends to be reduced to simple relocation, that is, the actual movement of people, whereafter they are largely left to fend for themselves.’ (4641). Bhan notes similar issues with a state-centric approach when displacement of people can be more internal to the city, especially when people in informal housing are moved from one settled part of the city to re-settle in another (Bhan note 5 above).

65 See also Shantistar Builders v Narayan Khimalal Totame, (1990) 1 SCC 520 (‘Shantistar Builders’). Here the members of poorer sections of the society challenged the permission granted by the Government of Maharashtra, exempting the land under the Urban Land Ceilings Act in the favour of a builder on the condition that he use it to make 17 000 tenements for the poorer section. The builders had sought to increase the price of these houses. In Mumbai the government sought to remove slums and allot the vacant land to private builders for construction of multi-storied tenements while allotting some of these apartments to families that lived in slums in that area. Gradually, the private builders were given further exemptions that allowed them to increase the prices on these houses given the ever-increasing demand for houses in Mumbai city. Weinstein & Ren (note 10 above) posit that critical social solidarities among citizens and civil society have ensured that the government was challenged whenever it sought to remove slum dwellers. Emphasising their scepticism with a state-centric approach, they contend that democratic participation ensured that the government was forced to provide alternate housing or was at least challenged in Mumbai, unlike Shanghai.

66 Ajay Maken (note 13 above) 52–53.
a brief discussion on the crisis of habitation in India and South Africa before proceeding with evaluating the extent to which there is constitutional recognition of the right to the city through housing rights.

C The crisis of habitation

The Supreme Court of India, observing the crisis of homelessness, noted that the ‘deshstitute in urban areas continue to suffer without shelters.’ Further, ‘[in spite] of the availability of funds and a clear mechanism through which to disburse them, we see an extremely unsatisfactory state of affairs on the ground.’ This is despite the ‘continuous monitoring of the matter’ by the court itself. The court went on to constitute a committee to provide recommendations to the central and state governments to ensure at least temporary shelters for the urban homeless. The 2011 Census of India puts the number of homeless people at 1.77 million people, while civil society experts believe that this is a gross underestimation. They peg the number at around three million. Consider that in 2017, over 1 600 people died in India due to exposure during extreme weather conditions. They lived not in ‘buildings or census houses’ but ‘in the open on roadside, pavements, in Hume pipes, under flyovers and staircases, or in the open in places of worship, mandaps, railway platforms, etc.’ The absence of a dedicated constitutional provision and moreover a national-level framework law effectuating housing rights has been a matter of concern. The Universal Periodic Report Review noted that the national urban housing shortage was projected at 34 million units in 2011. It notes the failures of policies in ensuring security of tenure, proper resettlement, prevention from forced eviction, and coping with frequently occurring disasters. The review noted with concern that housing deficits need to be tackled at a structural level, and the ‘smart cities’ project should not result in forced evictions and segregation. Yet, it was reported that in 2017 government authorities demolished over 53700 homes and evicted around 260 000 people for reasons such as ‘city beautification’, ‘slum-free city’, organizing mega-events, and ‘development’ projects. The report says that in most cases, ‘the state has not provided resettlement; where provided, resettlement is largely inadequate. Forced evictions are thus contributing to a rise in homelessness.’

67 ER Kumar v Union of India, W.P. (Civil) no. 55/2003 (Judgment delivered on 11 November 2016).
69 Ibid.
Unlike India, South Africa provides a constitutional guarantee in the form of a ‘right to access to adequate housing.’ Yet, the South African Human Rights Commission notes that ‘the country continues to face significant challenges in providing access to adequate housing to poor and vulnerable persons.’ The Census suggests that even though the percentage of people under ‘informal dwelling’ has declined over time, the absolute number has increased and stands in excess of two million. The South African Human Rights Commission laments that existing policies fail to adequately address the concerns of a variety of people, although ‘mechanisms are available for ensuring that even the most destitute of individuals are accommodated, their needs are not adequately addressed.’

While homelessness, as discussed above, is a critical concern in both places, Berger also draws attention to a much more widespread phenomenon he calls ‘housing indigency’, which concerns grossly unmet housing needs. Homelessness and housing indigency make people vulnerable not only to the elements of nature, they are also susceptible to health and nutrition risks. They lack access to the ‘basket of goods’ necessary to meet accepted standards of living that disables them from participating in social and political life of a city and, in that sense, they become disenfranchised. ‘When one can no longer inhabit a public space, have one’s possessions and shanty towns (home, by some definitions) burned or bulldozed, be arrested for one’s status rather than a crime (hence signalling a loss of civil rights), and only exercise political power with extreme difficulty, one cannot be said to be a citizen.’ Homeless persons are also rendered powerless as they are excluded from democratic decision-making processes of the city that produces such exclusionary laws and regulations. In that sense, they are merely the recipients of decisions with a rather limited voice in the politics that would ultimately determine their destinies. This powerlessness is not an individual problem, but a politico-economic one. It is felt in the politico-economic process underway in both India and South Africa, which has influenced not only the consumption habits of people, but also the spatial patterns of their cities. These countries are witnessing a burgeoning consumption-oriented

74 Section 26 of the Constitution of South Africa declares a universal right of access to adequate housing. It sets up a baseline against forcible evictions while it mandates the government to progressively realise this right through legislative and other measures.


77 SAHRC Report (note 75 above) 10.

78 Berger (note 53 above) 316–317. In fact, in 1990 the UN noted that while there are about 100 million homeless people worldwide, the population of those inadequately housed could be over a billion.

79 Given that there are overlapping causes for hunger and homelessness (including lack of affordable housing), unemployment and poverty, it is likely that they would tend to have similar socio-economic profiles as well. The WHO Health Principles of Housing (1989) lists eleven basic principles that establish a relationship between housing environment and health of inhabitants.

80 T Walsh & C Klease ‘Down and Out? Homelessness and Citizenship’ (2004) 10 Australian Journal of Human Rights 77, 80. The article uses TH Marshall’s citizenship theory. It keeps ‘equality of status’ at the centre to argue that homeless people, due to denial of material goods and access to social and political life (basket of goods), experience a sub-citizen existence. It departs from the formal-legal idea of citizenship to a more social-functional concept and includes all things that community members expect the government to secure – forming the basis of an implied social contract (Ibid 81).


83 Arnold (note 81 above) 2.
middle class and an expanding market. Even as the market is presented as a panacea for several socio-economic problems that have historically gripped these states, the obsessive focus on the market and middle class has rendered the people with unmet housing needs invisible within the dominant political culture and public spaces. Fernandes calls this ‘invisibilisation’ a result of deliberate political choices – the ‘politics of forgetting’. The politics of forgetting is manifested in laws that criminalise begging, squatting and panning, and regulations that spatially reorganise the city into zones and push the poor away to the peripheries. These state actions targeted at the poor render their basic subsistence and existence precarious without the authorities making any effort to reach out and engage with them. This exclusion and denial of agency, Arnold argues, has an identity dimension as well.

A clear example of the kind of politico-economic decision-making discussed above, is the recent trend of building ‘smart’ or ‘world class’ cities. This idea is presented as a new twenty-first century utopia that would integrate the urban with digital planning, and solve the problems of urbanization and sustainability. However, the SAHRC notes that this ‘world class city’ narrative has resulted in preference for private investments in the land situated close to the economic hub over the needs of the poor, thus shifting them to the outskirts of the city, removed from access to economic opportunities. In India, the ‘smart city’ plan is being canvassed with much vigour. It envisages significant foreign and private investments. It requires deliberation as to how this metamorphosis affects the dynamics and demographics of a city, and how the disadvantaged areas shall be included in the decision-making on the distribution of access to the resources of such cities.

To summarise, the claim of people with unmet housing needs, invoke issues of both economic and national identity manifested in an extreme form of marginalization and a feeling of uprootedness that renders them distant from the politico-economic mainstream. Moreover, the urban space is patterned by different social categories. The rich and poor neighbourhoods are located distinctly apart. Within a neighbourhood there are dominant races and castes. ‘The divisions in urban space are seen to both reflect and reinforce existing social and structural divisions in society.’ These social boundaries are objectified in social difference, manifested

85 Arnold (note 81 above) 2–4. Arnold argues that homeless persons are subject to stereotypes and ideological constructs. As a matter of economic identity, ‘homelessness represents the extreme case of this economic marginalization and thus is worth exploring for what it tells us about political economic norms, the status of democracy, and the deployment of prerogative power in the modern nation-state.’ (Ibid 3) As a matter of political identity, Arnold argues that homelessness is an experience of uprootedness. It results in such asymmetric power dynamics that homeless persons experience exclusion from the modern nation-state. (Ibid) The brute power of state that is unleashed on homeless persons through laws, zoning regulations and prerogatives for cleaning a city emphasises their ‘othering’ and disenfranchisement. (Ibid 6).
86 A Datta ‘New Urban Utopias of Post-Colonial India: Entrepreneurial Urbanization in Dholera Smart City, Gujarat’ (2015) 5 Dialogues in Human Geography 3. Analysing the disjuncture between the speed in investments and the persistent local protests, Datta argues that the fault-lines of Dholera, the first Indian smart city, are ‘built into its utopian imaginings, which priorities urbanization as a business model rather than a model of social justice.’ (emphasis supplied). This reflects prioritising of exchange value over the use-value of the city and a dilution of the social function of the city emphasised by the World Charter and several other documents and reports on the right to the city discussed above.
87 SAHRC Report (note 75 above) 9.
in unequal access to and distribution of resources, amenities and other usufructs of the city.\textsuperscript{89} The right to the city, therefore, is not only a redistributive demand for shelter and access to amenities, but alongside it, it is a substantive recognition-based claim for citizenship. This claim addresses not only the immediate loss, but is rooted firmly in the discursive processes of history, where the dominant groups through the legal and political apparatus have determined the location and access of the disadvantaged groups in and to the city.

D The right to the city and history

While Lefebvre saw industrialization and the advent of capitalism as the point of departure in his analysis of cityscape,\textsuperscript{90} the more recent works credit it to the rise of global capitalism and emerging discourses on modern, clean and smart cities.\textsuperscript{91} The turn from cities seen from their ‘use value’ to the instruments of ‘exchange value’ heralded by industrialization found impetus in the rise of global capitalism. Harvey notes that this shift has not only created new infrastructure, but also effected a radical transformation in lifestyles and a strong affinity for property rights.\textsuperscript{92} Illustrating the material social change that urbanization brought in South Africa, Alan Mabin notes that “[the] pressures of land loss, military exigency and a growing commercialization of exchange relationships rendered both individuals and whole communities susceptible to involvement in the growing wage-labour economy of the towns by the 1850s.”\textsuperscript{93}

Further, as the imperial state expanded, most people were deprived of independent control of what they saw as their land, even as they lived there, as they inevitably sent one or more of their family members to participate in the urban economy. The land, divided into reserved and non-reserved lands, saw the African population falling through the cracks and settle as squatters or tenants. While this phenomenon was observed most glaringly in the non-reserved parts, it was also visible in the reserved areas.\textsuperscript{94} Modernity brought ‘betterment’ of agricultural communities and exclusivity which created a large landless population that inevitably stayed in miserable settlements and sent their members to take part again in the urban economy. The inadequate urban policy and unfulfilled promise of housing, along with apartheid ideology, resulted in mass and brutal evictions and racialised zoning. These overcrowded settlements suffered from a shortage of resources, denial of access to healthcare and widespread unemployment. Further, laws like the Prevention of Illegal Squatting Act made the very existence of an African landless population a crime. Therefore, Dhiru Soni states that ‘[the] housing question in South Africa, especially for blacks, pervades their very existence: who they are, what they are, and where they

\textsuperscript{89} Ibid at 32.
\textsuperscript{90} Lefebvre (note 17 above) 66–85.
\textsuperscript{91} Harvey (note 14 above); Fernandes (note 84 above); A Sugranyes & C Mathivet (eds) Cities for All: Proposal and Experiences towards the Right to the City (2010).
\textsuperscript{92} Harvey (note 14 above) 319.
\textsuperscript{93} A Mabin Dispossession, Exploitation and Struggle: an Historical Overview of South African Urbanization in DM Smith (ed) The Apartheid City and Beyond: Urbanisation and Social Change in South Africa (1992) 13. Mabin notes that migration from villages to towns pre-dates colonization. However, with the growth and exploitation of the diamond industry presided over by imperial forces, the cityscape and social relations, particularly around Kimberley changed rapidly and drastically.
\textsuperscript{94} Ibid. 14–15.
stay… Housing, therefore, becomes an indicator and a potent symbol of the shifting power relations between classes and within different sectors of capital.95

Indian cities have historically been segregated along the lines of caste and religion, which were reinforced by colonialism.96 Despite being seen as emancipatory spaces of opportunities for Dalits, whose influx into the cities has grown by more than 40 per cent between 2005 and 2015,97 the urban localities tend to arrange themselves along caste and religious lines. Jodhka, in his field study on Dalit entrepreneurship notes that since most of them started with limited means, they ran small grocery shops which were located in Dalit residential areas. Since people in town knew about their Dalit origin, they would not provide them with space in their shops for rent, despite the constitutional promise of the abolition of untouchability.98 In fact, a relatively recent study shows that identity-based spatial exclusion is still rampant even in the big Indian cities.99 The trends of rising spatial exclusion is not limited to caste but also relates to religious minorities, particularly Muslims, in medium to large cities saddled with histories of communal violence.100 Discussing the exclusion of Muslim women from access to a city life, Sameera Khan says that this is closely linked to the exclusion of Muslims as a whole from the mainstream cultural, political and social fabric of Mumbai, including access to mixed housing.101 Further, she observes that while Muslims have historically lived in community-based enclaves in Mumbai, the communal riots of 1992–1993 pushed a sizeable number of them, particularly the poorer ones, who had lived in mixed housing to these enclaves. The spatial reorganization was followed by relative barring of Muslims from economic and business ties in the heterogeneous parts of the city. This experience has parallels with the experiences of Dalits as well. The ghettoization of minorities and vulnerable groups ranges from stark to subtle, but their ability to participate in the everyday production of the city and appropriate

95 D Soni ‘Apartheid State and Black Housing Struggle’ in DM Smith (ed.) The Apartheid City and Beyond: Urbanisation and Social Change in South Africa (1992) 52.
98 S Jodhka ‘Dalits in Business: Self Employed Scheduled Castes in North West India’, (2010) 45 Economic & Political Weekly 41, 43–44. The post-1991 economic policy entailed growth of private sector and the retreat of the state from most economic activities. While the Scheduled Castes (Dalits) benefited from quotas in public employment, there are no quotas in the private sector. There were Dalit entrepreneurs before 1991 too, but the new economic policy compelled them to look for jobs in the private sector or self-employ themselves to participate in the urban economy. Thus, significant numbers of Dalits migrated from rural areas towards the urban areas. However, the spatial distribution of residences manifested in caste-based ghettos and colonies. Migration also helped a small number of them to ‘hide’ their caste identities, yet where people knew of Dalit origins of a family, the identity itself became barrier to entrepreneurial opportunities in terms of finding a place to start a business.
100 Sahoo (note 97 above) 2.
101 S Khan ‘Negotiating the Mohalla: Exclusion, Identity and Muslim Women in Mumbai’ (2007) 42 Economic & Political Weekly 1527. Mohallas are the interior neighbourhoods of a city dominated by a community. The question Khan explores is whether living in these Mohallas dominated by their own community (Muslims) has a bearing on women’s spatial mobility. Drawing from an historical and ethnographical study, Khan invites us to look at intersections of class, gender and religion to have a more holistic understanding of this exclusion. The article also notes the role of 1992–1993 communal riots in ghettoization of Muslims. These patterns of social inequality mapped onto the spatial inequalities could be seen in other important cities of India too. See R Susewind ‘Muslims in Indian Cities: Degrees of Segregation and the Elusive Ghetto’ (2017) 49 Environment & Planning 1286.
its advantages is severely constrained, which reinforces the confluence of identity-related issues with issues of access and distributive justice. Besides the caste and religious identities, the economically weaker sections of society also reside in cluttered housing or slums, generally in the peripheral or the neglected interior of the city. Given the irregular and informal state of their housing, they are often subjected to evictions and displacement under the guise of law and planning. The residential clustering, whether in the form of ghettos, enclaves or slums, blends with another urban group because different ascribed and acquired identities may overlap: for example, religious minorities, backward castes (ascribed) and economically backward (acquired) identities could overlap. Jaffrelot and Gayer, studying the ghettoization of Muslims in India, outline the following characteristics to determine a ghetto: (i) element of social/political constraint in the housing options of a given population, (ii) class and caste diversity which regroups individuals on the basis of ascribed identities, (iii) the neglect of these localities by the state authorities; (iv) estrangement of these localities and their residents from the rest of the city due to lack of transport, jobs and access to public spaces; and (v) a subjective sense of closure these residents feel from the objective patterns of estrangement from the city, to study the residential clustering in major cities of India. In crucial ways, this spatial relegation, therefore, is both an outcome as well as a cause for the denial of the right to participate in and the right to appropriate the advantages of the city.

The experiences of both nations suggest that unequal distribution of housing, transport, water, sewage and such amenities necessary to access the city is not residual, but is produced through state actions. Heller and Mukopadhyay, writing in the Indian context, list several state actions such as deliberate under-planning of residential spaces that creates extreme dependence on local political power-wielders. As a result, provision of basic housing and attached amenities to access the city is ad hoc and subject to significant transaction costs. The public investments at the cost of demolition of several such unauthorised colonies often benefit the richer population through investment in specific transportation accompanied with the neglect of others. Thus, ‘the state first crafts a legal and regulatory framework to impact the poor disproportionately; it then builds on this framework to restrict quality service delivery to a small group of citizens and finally, it perpetuates these differences by making investments that primarily benefit the better-off.’ Heller and Mukopadhyay believe that this represents not only a denial of basic human capabilities, but importantly, the negotiated existence has a deleterious effect on local democracy; and the gross spatial inequality in delivery of basic entitlements may possibly cause the excluded settlements to harden into ghettos.

102 Susewind argues that it is not only the place of residence, but multi-level structural processes whereby people are selected, thrust and maintained in these locations, and the culture they develop therein, that is critical for studying urban marginality. This suggests the inextricability of history and sociology for studying the right to the city. (Ibid 1288–1289).

103 As quoted in, Susewind (note 101 above) 1289–1290.

104 Heller & Mukopadhyay (note 96 above) 52. While the authors restrict their study to Delhi, there are reasons to believe that their findings would be true for many other medium to large cities in India.


106 Ibid at 54.

107 Ibid. Additionally, the socio-economic profiling of such colonies and communities would suggest that they primarily house the historically underserved population groups including Dalits, Muslims and migrant labourers.
Part II has sought to provide a broad understanding of the right to the city and has suggested why the issue of homelessness and housing indigency could be understood from a wider right to the city framework. It has also argued that an inherently dynamic notion of the right to the city would study the sociological and historical process which uncovers the role of the state in marginalization of disadvantaged groups to the peripheries of the city, denying them access to amenities, opportunities and advantages offered by a city. Part III will trace the housing rights jurisprudence of South Africa and India, and analyse them from the perspective of the right to the city.

III  FROM HOUSING TO CITY: CONSTITUTIONAL RECOGNITION OF A RIGHT

Though neither country recognises the right to the city in their constitutions, both nations engage in robust constitutional jurisprudence in the domain of civil-political and socio-economic rights that provide important steps towards recognition of this right.

A  South African position

The South African Constitution contains a range of human rights, including civil and political rights, socio-economic rights and collective rights, partly aimed at correcting the wrongs of apartheid. The Constitution also enshrines the right to property. However, this right can be restricted for land reforms and redistribution to correct past racial discrimination. The Constitution holds the values of dignity, equality and non-discrimination to be central. If the right to the city is considered an ensemble of various rights that would enable a resident to enjoy access to various amenities of the city in a non-discriminatory manner, the Constitution has the resources effectively to recognise it. The right to the city has other features as well which could be recognised within the South African framework, such as democratic participation, use of public transport, priority of the use value of a city over its exchange value, and regarding the city as a common good.

The South African constitutional jurisprudence appears to adopt a more sympathetic approach to the right to housing in comparison to India. For example, in the Grootboom case, the Constitutional Court noted that the causes of unmet housing needs that compelled people to occupy land lay in the apartheid regime. Noting the state complicity, the Court described the way in which the state had engendered an ‘apartheid cycle’, which ‘was one of untenable restrictions on the movement of African people into urban areas, the inexorable tide of the rural poor to the cities, inadequate housing, resultant overcrowding, mushrooming squatter settlements, constant harassment by officials and intermittent forced removals.’ Inevitably, people had to live in appalling conditions on the encroached lands. The post-apartheid state too failed to provide them with reasonable housing despite being waitlisted for seven years. Hence, people chose self-help over waiting for the allocated low-cost housing by the state.

---

Section 26(2) of the Constitution requires the government to ‘take reasonable legislative and other measures, within its available resources, to achieve the progressive realization of this right [to housing].’ The *Grootboom* case, based on the right to housing, was significant from a norm-making perspective. The Court disapproved self-help over state provision, yet it was sympathetic to the situation of the occupiers. The judgment emphasised the indivisibility and inter-relatedness of rights and interpreted the constitutional text by keeping the context central. Further, it explained the test of reasonableness to include an evaluation of the validity of state action such that ‘housing problems [are considered] in their social, economic and historical context and to consider the capacity of institutions responsible for implementing the programme.’ The Court proceeded to analyse the housing policy from this reasonableness perspective and held that while considering reasonableness, it ‘will not enquire whether other more desirable or favourable measures could have been adopted, or whether public money could have been better spent.’ This perspective accepts a range of measures are possible to meet the constitutional obligations. ‘The programme must be balanced and flexible’ and ‘[should] attend to housing crises and to short, medium and long term needs.’ It would be *ex-facie* unreasonable if the programme ignores ‘[those] whose needs are the most urgent and whose ability to enjoy all rights therefore is most in peril’. The programme must be context-sensitive and subject to continuous review and the availability of resources should be considered in determining the reasonableness of any programme.

While the Court found the government’s programme reasonable with respect to attending to medium and long term needs, it declared the policy inadequate because ‘no provision was made for relief to the categories of people in desperate need.’ Further, it emphasised a humane approach towards eviction, particularly for people with desperate needs. The Court admonished the agencies for the inhumanity with which the evictions were carried out where the residents’ possessions were not only removed but burnt and destroyed, and held that an eviction programme must uphold the values of equality and dignity. Irene Grootboom, however, died destitute and homeless despite a favourable judgment. This suggests that a favourable judgment may not translate into reality if it is not adequately followed up. Yet, the judgment has some important outcomes for developing jurisprudence on the right to the city. The judgment suggests that balanced regional growth, especially in the rural areas, is necessary to arrest uncontrollable migration into urban areas and prevent housing crisis. Further, it noted state complicity in allowing the informal settlement in New Rust to swell. Importantly, the Court extended the scope of the right to housing from merely shelter to also include provision of essential services.

---

111 Emphasis added. The similar language of ‘reasonable measures’ can be seen in s 25(5) (access to land); s 27 (right of access to healthcare, food, water and social security); Section 29 (right to basic education).
112 *Grootboom* (note 110 above) paras 59, 80–81.
113 Ibid at para 43.
114 Ibid at para 41.
115 Ibid at para 43.
116 Ibid at para 44.
117 Ibid at paras 64–65.
118 *Port Elizabeth Municipality v Various Occupiers* [2004] ZACC 7, 2005 (1) SA 217 (CC)(‘*Port Elizabeth Municipality*’)(Sachs J emphasises the historical shift from Prevention of Illegal Squatting Act No. 52 of 1951 to the Prevention of Illegal Eviction and Unlawful Occupation of Land Act No. 19 of 1998 (PIE Act)).
Another feature of the South African jurisprudence is its emphasis on the social purpose and the use value of property in the balancing property rights with other interests of distributive and corrective justice. In the Port Elizabeth Municipality case, 68 people, including 23 children, had occupied 23 shacks that they had erected on privately owned, albeit undeveloped, land for several years. Upon receiving a petition signed by 1,600 people, the Municipality ordered their eviction. This situation pitted the property rights of a few against the basic necessity needs of many. Interestingly, the occupiers had acceded to leave the land if they were given alternative housing and reasonable notice. When it was proposed that they be relocated to another township, Walmer, they refused to go as Walmer was unsavoury, over-crowded and crime-ridden. Since the Municipality was already working on a housing development programme, it argued that providing an alternative space to the occupiers would amount to ‘queue-jumping’.

The Court responded that the motivation of the occupiers (whether they intended to jump the queue), the duration for which they had been staying and the existence of alternative arrangements (among other factors) should be considered before passing eviction orders. Accordingly, it called for agencies to be ‘far more cautious in evicting well-settled families with strong local ties, than persons who have recently moved on to land and erected their shelters there.’ The Court added another layer of security by mandating that people must not be punished for not complying with eviction orders unless the eviction order is also a judicial one. Margaret Radin categorises private property in terms of its relationship with its owner/possessor into personal and fungible. While personal property is one that is ‘bound up’ with the person, fungible property is held only instrumentally. Therefore, when a person has resided on a property for sufficiently long, one is considered ‘entrenched’ in that property. The well-settled families with local ties have invested their personhood in that informal housing as it bonded with their personhood. Those who moved relatively recently, on the other hand, are not similarly entrenched. Hence, even though their shacks are important to them, the law could regard them as fungible homes. With the effluxion of time, a fungible item could become personal without changing hands, as ‘people and things become intertwined gradually.’ It is argued that the entrenchment of a well-settled family is not confined to their individual self and dwelling but through local ties and navigation; these are entrenched in the city too. In line with the classification maintained in this article, the personal tends to impact on the use value,

---

119 Ibid at para 15. The PIE Act, Sachs J explains recognises this balance as well as difference in eviction proceedings brought by a private land-owner and municipality.
120 This argument seems to be consonant with the reasoning of the Court in the Grootboom case (note 110 above).
121 Ibid at para 27.
122 Ibid at paras 49–50.
123 M Radin ‘Property and Personhood’ (1982) 34 Stanford Law Review 957. It is not suggested that ‘fungible property is unrelated to personhood, but simply that distinctions are sometimes warranted depending upon the character and strength of connection.’ For example, a home is generally personal, while vacant plots appear more fungible. One reason could be necessity, but it is also true that one feels her home as a place for self-expression, comfort and security. Dispossessing someone of a ‘personal’ property calls for greater caution, as it attacks subject’s person and dignity.
while *fungible* lies in the realm of exchange-value of property and city. The Court noted that neither the owner nor the Municipality needed the land for any immediate productive use (or, in Radin’s terms, relatively fungible). Yet, given that the government is expected to both ensure that an owner’s property rights are protected and the basic needs of the occupiers are met, the Court required all the interested parties and the Municipality to meaningfully engage and arrive at an acceptable solution.

The rule of ‘meaningful engagement’ was a judicial innovation from the Court in the case of the Olivia Road occupiers. In *Olivia Road*, the City of Johannesburg sought to evict a number of impoverished occupiers of the buildings that were deemed structurally unsafe. The City argued that the occupation constituted a threat to the health and safety of the occupiers themselves. The Court passed an interim order directing the City and occupiers to ‘engage with each other meaningfully… in an effort to resolve the differences and difficulties aired in this application in the light of the values of the Constitution, the constitutional and statutory duties of the municipality and the rights and duties of the citizens concerned.’ The Court ordered the parties to report the results of the engagement, the substance of which shall be considered in further orders or judgment. The parties arrived at a comprehensive settlement which made their eviction conditional on providing alternative accommodation, which in turn was held conditional on providing a suitable permanent housing solution in consultation with the occupiers concerned. Liebenberg argues that, while the ‘meaningful engagement’ order exhibited key elements of deliberative democracy, which facilitated a participatory and contextualised solution to the problem of safety and pressing need for shelter for many people, the Court failed to answer the more systemic question raised by the occupiers – whether the City had put in place a reasonable plan for the permanent housing of the occupiers and other similarly situated 69 000 inhabitants of the inner city. The Court hoped (like

---

125 The personal/fungible classification is not always apposite. For people who recently moved into a place, may not have done so with the intent of mere instrumental use of the property. Yet, in the context of the right to the city, ‘entrenchment’, a factor of time, could itself suggest a presumption for treating older settlements as personal as opposed to newer shacks. Radin says, ‘A person cannot be fully a person without a sense of continuity of self over time. To maintain that sense of continuity … one must have an ongoing relationship with the external environment, consisting of both “things” and other people.’ (Ibid 1004).

126 [*Port Elizabeth Municipality* (note 118 above) at para 59.

127 *Occupiers of 51 Olivia Road Berea Township and 197 Main Street Johannesburg v City of Johannesburg & Others* [2008] ZACC 1, 2008 (3) SA 208 (CC) (‘*Olivia Road*’). (The Court held that ‘larger the number of people potentially to be affected by eviction, the greater the need for structured, consistent and careful engagement.’ The engagement must be in good faith and ‘people in need of housing must not be regarded as a disempowered mass.’ (Para 20). In *Olivia Road* the Constitutional Court rejected the view of the Supreme Court of Appeals that the duty to act on the part of the City and the right to housing of the occupiers of unsafe buildings are not reciprocal and that these cases only peripherally raise the questions of constitutional obligations of the state organs. The concept of ‘meaningful engagement’ developed in this case, respects people’s right to participate in the decisions pertaining to their relation with their housing, neighbourhood and rehabilitation, prior to eviction. Contrast this with a minimalist and formal notion of hearing seen in the judgments of the Indian Supreme Court. In more recent times the Supreme Court has, by interpreting them as licencees, put further limitations on the rights of the inhabitants. U Ramanathan *Illegality and the Urban Poor* (2006) 41 *Economic & Political Weekly* 3193, 3195.

128 *Olivia Road* (note 127 above) para 5.


130 Ibid at 18.
Olga Tellis\textsuperscript{131} in India, discussed below) that the City would carry out the consultation in good faith. Given the vague rules of engagement, Liebenberg says that ‘there is a real danger that meaningful engagement as an adjudicatory strategy may descend into an unprincipled, normatively empty process of local dispute settlement’ if appropriate regulatory measures and resource allocations are not undertaken.\textsuperscript{132}

In the case of the Joe Slovo evictions,\textsuperscript{133} for example, even as the Court assumed greater control by fixing baselines for alternative housing and obligating the parties to negotiate the details of the time and circumstances of relocation, the Court effectively condoned the top-down heavy-handed negotiation carried out by the government.\textsuperscript{134} Joe Slovo, like Shantistar Builders\textsuperscript{135} in India (discussed later), is a tale of broken promises. When the residents in Joe Slovo were asked to relocate to Delft to facilitate an \textit{in situ} major housing development, they were promised that 70 per cent of those relocated would be given low-cost housing in Joe Slovo itself. The first phase of the project did not give effect to the promise and subsequently the rentals were pitched far higher than initially envisaged. The emphasis was placed on ‘bonded housing’ ie, market rate for housing purchased through a mortgage. Even though the Court determined the baselines for alternate housing, ordered relocation of 70 per cent of evictees and directed the parties to engage on the matters of detail, the order should not be regarded as a substitute for a well-structured and regulated process of engagement laid down in law. Notably, in Joe Slovo, the Court also said that the utilitarian gains of housing development projects could outweigh the minor defects of engagement. This treatment of ‘meaningful engagement’ is quite different from Olivia Road, which conceptualised it as a mechanism that (a) treated people facing evictions as active participants rather than ‘disempowered mass’, and (b) respected the democratic mandate and the institutional expertise of the legislature and the executive.\textsuperscript{136} With Joe Slovo, it seems that the role of ‘meaningful engagement’ has been limited to only procedural compliance which could be subservient to other factors in determining the validity of state action,\textsuperscript{137} even though it was clear that the process was flawed in terms of lack of people’s participation in formulation of the scheme. Dugard, in fact, opines that by avoiding the questions of location in the alternative accommodation plans and the failure to provide for housing for poor in the inner city housing plans in determining the validity of the City’s

\textsuperscript{131} Olga Tellis v Bombay Municipal Corporation, (1985) 3 SCC 545 (‘Olga Tellis’).
\textsuperscript{132} Ibid at 19. For example, in Mamba v Minister of Social Development CCT 65/08, the government sought to dismantle the refugee camps built in the wake of xenophobic violence. The Government interpreted meaningful engagement minimally, though ordered in the same terms as Olivia Road (note 127 above), and went ahead with the dismantling the settlement. This case emphasises the need for a structured long term process of engagement. See, B Ray ‘Engagement’s Possibilities and Limits as a Socioeconomic Rights Remedy, (2010) 9 Washington University Global Studies Law Review 399, 404–408.
\textsuperscript{133} Residents of Joe Slovo Community, Western Cape v Thubelisha Homes, [2009] ZACC 16, 2010 (3) SA 454 (CC).
\textsuperscript{135} Shantistar Builders (note 65 above).
\textsuperscript{137} There seems to be confusion among judges whether the engagement in Joe Slovo amounted to what O’Regan J described as ‘meaningful engagement’. The question for her was whether absence of such engagement was enough to deem the project as unreasonable. Sachs J. believed that there was meaningful engagement but that is only one of the factors to be considered to determine the reasonableness of state action. (Ibid 744–745).
housing policy, the Court side-stepped the core issues in *Olivia Road*. While, ‘meaningful engagement’ is a welcome innovation, Dugard believes ‘[it] does not provide poor people with any concrete protections against eviction, nor does it help to delineate the right to housing’.

In another case, engaging with the rhetoric of creating clean and orderly cities to justify the eviction of informal traders, the Court held that while these objectives are laudable, they cannot be achieved by ‘flagrant disregard’ for the rights of people. Yet, it restricted the application of its judgment to only the lawful occupiers of the space and allowed the City to ‘use all lawful means to combat illegal trading and other criminal conduct’ even though this could severely impact other inter-related constitutional rights such as right to dignity, freedom of trade and socio-economic rights of children. Pieterse notes that the Court, with its eye on only the immediate issues of legal compliance, side-stepped more profound contemplation of the constitutional rights of traders. A ‘right to the city’ based inquiry may have been helpful here in situating the inter-linked constitutional rights of trade, livelihood, housing and the rights of dependent members of family, in the broader framework of citizenship claims animated by notions of access and participation in the production of the urban space. Pieterse notes that due to its narrow focus on ordinary legal compliance, the Court order ‘at best fails to disturb and at worst insulates the manner in which the notion of legality itself contributes to the marginalisation and exclusion of vulnerable residents in South Africa’s post-apartheid cities.’

Interestingly, in *Yolanda Daniels* the Court abandoned the narrow focus on ordinary legal compliance and drew on the histories of dispossession and confinement in colonial and apartheid regimes to draw linkages between recognising the legal security of tenure and the value of human dignity cherished by the Constitution. The study of historical context then informed the purposive interpretation of Extension of Security of Tenure Act 62 of 1997 (ESTA) to rule that an occupier has the right to carry out necessary improvements to make the shelter habitable. De-emphasising the private property rights, the Court held that it was empowered to impose positive obligations on a private person vis-à-vis the occupier. It considered the nature, history and the purpose of the right, and whether adopting a hands-off approach vis-à-vis a private person would effectively negate that right. Lastly, the Court

---


139 Ibid 238.


141 Ibid. For the historical account of negotiated existence of informal traders and City administration, see M Pieterse Rights, Regulations and Bureaucratic Impact: The Impact of Human Rights Litigation on the Regulation of Informal Trade in Johannesburg, available at http://www.scielo.org.za/pdf/pelj/v20n1/04.pdf. Pieterse critiques the false dichotomy between legality and illegality in informal street trade adopted by earlier policies of the City; and thereafter a complete erosion of the distinction when the City found irregularities in the grant of permits under earlier policies such that every informal trader, including the permit-holders, was removed under ‘Operation Clean Sweep’. He criticises the assumptions behind equating the informal traders with dirt to be eradicated from the city.

142 Pieterse (note 141 above) 13–14. Pieterse notes that the order of the Court did not stop the blitz approach of the City but it appeared to be more careful in ensuring that it acted within the bounds of its by-laws. The eviction of informal traders on the intersections, for example, went unchallenged as it complied with the ordinary laws of the City.

143 Ibid 22–23.

144 Daniels v Scribante [2017] ZACC 13, 2017 (4) SA 341 (CC) (‘Yolanda Daniels’).

145 Ibid at paras 14–22.

146 Ibid at para 39.
suggested that meaningful engagement could help balance the owner’s rights over the property and those of the occupier.\textsuperscript{147}

The Court interpreted the right to housing by keeping in view the socio-historical contexts of the housing indigency. It was prepared to give a more expansive meaning to housing than merely shelter and called for a more humane and dialogic approach to eviction keeping the values of dignity and equality at the centre. Further, the Court, through ‘meaningful engagement’ has entrenched the need for stakeholder participation in deciding on eviction, the timelines, as well as rehabilitation. The Court has also been prepared to de-emphasise the private property rights of a few in the interest of the use rights of large groups of impoverished people. Yet, the lack of clarity on the substantive content and absence of regulatory safeguards threaten to rob ‘meaningful engagement’ of its purpose to ensure people participate on the decisions of their (dis)location in their City.

\textbf{B Indian position}

Contrary to the South African Constitution, the Indian Constitution has most of the socio-economic rights in its non-justiciable Part IV called the ‘Directive Principles of the State Policy’. However, the underlying objective of the Constitution is to give effect to a social revolution in India. Post 1978, the judiciary also emerged as an equal stakeholder in this revolution. The Supreme Court of India through a series of judgments relaxed the \textit{locus standi} rules and democratised the access to justice through Public Interest Litigations (PIL). It interpreted fundamental rights liberally to strengthen anti-discrimination law and underscore the indivisibility of civil-political and socio-economic rights while providing more substance to the ‘right to life’ enshrined in Article 21 of the Constitution.\textsuperscript{148} In \textit{Francis Coraile Mullin}, Bhagwati J. famously observed:\textsuperscript{149}

By the term ‘life’ as here used something more is meant than mere animal existence….the right to life includes the right to live with human dignity and all that goes along with it, namely, the bare necessities of life such as adequate nutrition, clothing and shelter and facilities for reading, writing and expressing one-self in diverse forms, freely moving about and mixing and commingling with fellow human beings.

The Supreme Court then proceeded to read the right to elementary education,\textsuperscript{150} health,\textsuperscript{151} food, clean environment, shelter and livelihood, as well as access to justice\textsuperscript{152} within the conspectus of the right to life. The harmony between the directive principles and fundamental rights has also been held to be a part of the unalterable basic structure of the Constitution.\textsuperscript{153}

\textsuperscript{147} Ibid at para 62.

\textsuperscript{148} For a comprehensive overview of judicial activism in India, see S Sathe ‘Judicial Activism: The Indian Experience’ (2001) 6 Washington University Journal of Law \\& Policy 29, 43–63.

\textsuperscript{149} \textit{Francis Coraile Mullin v The Administrator, Union Territory of Delhi}, (1981) 1 SCC 608, 618–619.

\textsuperscript{150} \textit{Unnikrishnan v State of Andhra Pradesh}, (1993) 1 SCC 645.


\textsuperscript{152} \textit{Hussainara Khatoon v State of Bihar}, (1980) 1 SCC 81.

\textsuperscript{153} \textit{Minerva Mills v Union of India}, AIR 1980 SC 1789.
The basic structure doctrine, developed by the Supreme Court,\textsuperscript{154} holds that the Constitution has certain essential features which cannot be amended. These features are not contained in specific provisions but pervade its overall scheme (e.g., rule of law, democratic governance, independence of judiciary, separation of powers, federalism, secularism, etc.). By interpreting the harmony between the directive principles and fundamental rights as part of the basic structure, the Court reinforced the indivisibility of and interdependence between the civil and political rights and the socio-economic rights.

In the \textit{Olga Tellis} case, a PIL was filed after the Maharashtra Government and the Bombay Municipal Corporation (BMC) decided that slum and pavement dwellers who lived in various parts of the city in deplorable conditions should be forcibly evicted. They were to be either deported to their place of origin or removed to places on the exterior of Mumbai. Pursuing that order, some pavement dwellings had been demolished when the petition was filed before the Supreme Court. The petitioners contended: (i) evicting a pavement dweller from his \textit{habitat}\textsuperscript{155} amounts to depriving him or her of their right to a livelihood which is guaranteed by Article 21; (ii) the order violates their right to residence and occupation covered under Article 19; (iii) the procedure under the BMC regulations regarding eviction as arbitrary inasmuch as it dispenses with any duty to provide notice; and (iv) it violates the constitutional spirit to classify the pavement dwellers as ‘trespassers’ because they have been forced to take up inadequate housing due to economic compulsions. The petitioners requested the court to determine the scope of the concept of ‘property’ in a welfare state given the constitutional mandate that property should serve the common good.

The court, recognising the harmony of directive principles and fundamental rights, observed that in light of the constitutional mandates for securing the common good, equitable distribution of wealth and material resources, and adequate means of livelihood,\textsuperscript{156} it was important to read livelihood within the scope of Article 21. Drawing linkages among the rights to livelihood, shelter and life, the court held that the pavement dwellers ‘choose a pavement or a slum in the vicinity of their place of work, the time otherwise taken in commuting and its cost being forbidding for their slender means. To lose the pavement or the slum is to lose the job… [and consequently] the deprivation of life.’\textsuperscript{157} The court, stopping short of striking the BMC regulation down, held that even if the regulation authorised the Commissioner to dispense with notice, it should be read as an exception and not the rule and only in exceptional circumstances could the Commissioner dispense with notice. Curiously, the court felt that the opportunity granted to the petitioners to make their arguments before the court effaced any need for a post-decisional hearing from the Commissioner. On the question of property rights, the court held that encroaching public property is an unauthorised use and technically, a ‘trespass’. Yet, given the circumstances of helplessness in which the poor people end up residing in such informal clusters, it does not amount to a criminal act. The court directed

\textsuperscript{154} Kesavananda Bharati v State of Kerala, (1973) 4 SCC 225. (This was an historic case decided on April 24, 1973. 13 judges of the Supreme Court decided on the constitutionality of certain constitutional amendments and on the question of limits on the amending powers of the Parliament. By a majority of 7:6, the Court held that while Parliament could amend any part of the Constitution, it cannot amend certain basic features of the constitution such as democratic form of government, rule of law, separation of powers, federalism, etc.).

\textsuperscript{155} Emphasis added. It is interesting that the petitioners, in their arguments use the word ‘habitat’ as it conveys a wider import than merely a static residence.

\textsuperscript{156} Constitution of the Republic of India, 1950 Article 39.

\textsuperscript{157} Olga Tellis (note 131 above) 575.
however, that based on the undertaking of the government, the slum and pavement dwellers identified before 1976 should be resettled in a nearby suburb, without making it a condition for eviction. Those households that had been staying in slums for a long time and had carried out significant improvements should not be evicted unless their land was to be acquired for public purposes, in which case they should be properly re-settled. Ultimately the court directed the BMC to allow the pavement dwellers to stay in their shacks till the month after the end of the monsoon season.

The broad significance of this case lies in the recognition of the right against summary eviction and a legitimate claim for resettlement. Interestingly, the petitioners had also led evidence to establish that the burgeoning slums and pavement dwellings in Mumbai are a direct result of unplanned urbanization and state complicity. Yet, Olga Tellis offers scant support for the right to housing. In fact, the court explicitly held that the state is under no positive obligation to provide shelter and livelihood to people. By not striking down the regulation, the court accepted that dwellings could be demolished without notice and people may be evicted without a commitment for resettlement. The judgment effectively renders the right to housing conditional. It can often be used to justify eviction if formal compliance with the requirements of due notice are met.

In Shantistar Builders, the petitioners were living in slums on a public land. They were removed from there on condition that the government had tenements built on the same land and sold them at low prices to the families who were evicted. When the state government permitted the builders to increase the prices of these tenements in the exercise of its extra-ordinary powers under land ceiling laws, the petitioners challenged the permission. The Supreme Court observed that ‘[with] the increase of population and the shift of the rural masses to urban areas over the decades the ratio of poor people without houses in the urban areas has rapidly increased.’ Further, it said, ‘[in] recent years on account of erosion of the value of the rupee, rampant prevalence of black money and dearth of urban land, the value of such land has gone up sky-high. It has become impossible for any member of the weaker sections to have residential accommodation anywhere and much less in urban areas.’ The court held that ‘[because] reasonable residence is an indispensable necessity ... [it is] included in ‘life’ in Article 21.’ The exemptions provided by a beneficial law such as the ceiling law that should be monitored to ensure that the law was exercised for the benefit of the poorer

---

158 However, see Narmada Bachao Andolan v Union of India, (2000) 10 SCC 664.
159 Pillay (note 109 above) at 390–391.
160 Ibid at 391.
161 Ibid.
162 M Khosla ‘Making Social Rights Conditional: Lessons from India’ (2010) 10 I-CON 739. Khosla notes, ‘Olga Tellis provides for no individualised right to shelter, as well as no right that the state take reasonable measures to provide for shelter. The focus is on ensuring that a proper procedure for eviction is followed.’ (Ibid 747) ‘[The] Court’s focus was not on how a person’s right would have been satisfied if the state had not acted.’ (749). Following, Olga Tellis (note 131 above) the Supreme Court in Ahmedabad Municipal Corporation v Nawab Khan, (1997) 11 SCC 121 ordered that people residing in temporary hutments in certain streets of Ahmedabad could be removed after a due notice, if they were not eligible beneficiaries of existing government schemes. However, the Delhi High Court in Ajay Maken (note 13 above) considered Olga Tellis as precedent for the right to housing and livelihood.
163 Ibid at 527–528.
164 Ibid at 529.
165 Ibid.
communities. Even here, the court gave a restricted interpretation of the right to housing by only providing for ‘reasonable housing’. Importantly, it did not obligate the state to provide alternative accommodation in the event of forcible eviction.

As the Indian Government embraced the new economic policy of liberalization and privatization, the judiciary also changed its notion of ‘public interest’ and its perspective on slums and slum-dwellers. The PIL which allowed procedural departures in the interest of justice, now functioned almost like a ‘slum-demolition machine’. With the rising middle class, consumerism and institutionalisation of these ethics and aesthetics in creating ‘world class’ cities, the urban poor were now seen as ‘trespassers’ and slums were seen as ‘dirty’, ‘out of place’, a ‘nuisance’ and ‘large areas of public land usurped for private use free of cost.’ Bhuwania argues that ‘it is the PIL with the kind of power it vests in judges which actually enables them to act on such biases and that too with a free hand in a most expansive manner, unconstrained by technicalities and rules of adjudication, and on such flimsy evidence as random photographs’. These PILs were deployed to order a city-level clean-up of Delhi’s street vendors, beggars and cycle-rickshaw drivers’ but principally the city’s slums were at the receiving end of the court’s ‘wrath’ which ousted over a million people from the city. In the Supreme Court, such biases found their way into a PIL against the municipal garbage waste disposal practices in Indian cities, when Justice Kirpal turned the case towards an unrelated issue of slums in Delhi, blaming them for the solid-waste management problem of the city. It resulted in an infamous outburst by Justice Kirpal in the Almitra Patel case:

Establishment or creating of slums, it seems, appears to be good business and is well organised... The promise of free land, at the taxpayers cost, in place of a jhuggi [shack], is a proposal which attracts more land grabbers. Rewarding an encroacher on public land with free alternate site is like giving a reward to a pickpocket.

In the same judgment, the court ordered the Delhi Government to provide land for disposing of the city waste to the Municipal Corporation for free. This judgment marks a moment where (i) illegality was singled out as a trait of a slum-dweller and (ii) cleaning-up of the city and making it world-class assumed priority. This obfuscates the fact of continuous under-provisioning of houses for over four million slum dwellers of Delhi. At the turn of the twenty-first century, the Planning Commission noted that about 90 per cent of the shortage

166 Ramanathan (note 127 above) 3194.
168 A Ghertner ‘Nuisance Talk and Propriety of Property: Middle Class Discourses on Slum-Free Delhi, (2011) Antipode 1, 2, 7–8. Ghertner argues that ‘nuisance talk’ depicts slums as illegal environments. It reworked the public/private divide that inserts the codes of civility once restricted to home or neighbourhood into the core of public life. Further, it justified the privatization of public lands and urban restructuring (at 3). ‘While couched in the language of danger and insalubrity, … nuisance talk often betrayed more of a concern with property value and the quality of [private life]’ (at 8). For example, the rich people’s assertion of retaining the ‘posh’ value of their colonies by opposing densification and slum clusters, imports the private property concepts of ‘exclusivity’ into the public realm.
170 Bhuwania (note 167 above) 69–70.
171 Ibid. 71–72.
172 Almitra Patel (note 169 above) 685.
173 Ramanathan (note 127 above) 3195.
of urban housing in India pertained to urban poor and was attributable to non-provisioning of houses to slum-dwellers. Meanwhile, a number of structures that were constructed by the state agencies that had earlier not been provided for in the Master Plan of Delhi (a document mapping planned development of Delhi and a reference point for un/authorised constructions in the region), a procedural breach was identified and corrected. Some of these constructions replaced the existing slums, for example, on the banks of the Yamuna River.\(^{174}\)

In *Olga Tellis*, the petitioner argued that an important reason for the emergence of slums in metropolitan cities like Mumbai has been that the master plans of these cities have not been followed. The population was unevenly distributed because of severely imbalanced regional distribution of job opportunities. Consequently, poor people keep coming back to the city even after they were evicted.\(^{175}\) Yet, the apex court neither strengthened the rule of notice, nor made it mandatory for the state to provide alternative housing as a condition precedent for eviction. At the turn of the millennium, the courts regularly directed the government to carry out eviction as expeditiously as possible without needing to provide any substitute accommodation and due process.\(^{176}\) The High Court of Delhi in one petition identified the ‘public interest’ involved in the safety and hygiene concerns of the residents. The court said:\(^{177}\)

> After all, these residential colonies were developed first. The slums have been created afterwards which is the cause of nuisance and brooding [sic] ground of so many ills. The welfare, health, maintenance of law and order, safety and sanitation of these residents cannot be sacrificed ... in the name of social justice to the slum dwellers.

A decade later, the High Court extended the argument further and held that ‘the rights of the honest citizens... cannot be made subservient to the rights of encroachers’, which prioritises protection of private property rights as a matter of public interest.\(^{178}\) This makes the recent judgment in *Ajay Maken* remarkable.\(^{179}\) This was a PIL where the railway officials (the land where 5,000 slum-dwellers had been residing for over two decades was held by the government agency, Railways), and police had demolished the slums to forcibly remove the residents, leaving them without any shelter in extreme weather. Disregarding the rules laid down in *Sudama Singh*,\(^{180}\) the Railways had neither surveyed the location nor provided any advance notice to the residents. The Delhi High Court called the demolition drive a ‘human tragedy’ and observed that ‘[the] right to housing is a bundle of rights not limited to a bare shelter over one’s head. It included the other rights to life viz. the right to livelihood, right to

---

174 Ibid.
175 *Olga Tellis* (note 131 above) 563–564.
176 Ibid, 3195–3196. Space precludes discussion of each of these judgments of the Delhi and other High Court as well as Supreme Court. Bhan (note 5 above), Bhuwania (note 167 above) and Pillay (note 109 above) critique several judgments of High Courts and the Supreme Court that presided over forcible evictions and demolition of slums.
178 Ghertner (note 168 above) 18–19.
179 However, it also highlights a discontinuous and rather unprincipled approach to judicial activism in India. Pillay (note 109 above) suggests that genuine efforts towards engaging with the executive by the South African Constitutional Court in the enforcement of the socio-economic rights is a model worth considering.
180 *Sudama Singh v Government of Delhi*, 2010 (168) DLT 218 (‘Sudama Singh’). Given the anti-slum trend of judgments from the High Court of Delhi (See, Bhuwania (note 167 above), Bhan (note 5 above) and Ghertner (note 168 above)) the human-rights oriented approach of *Sudama Singh* seemed an outlier. The High Court, by following *Sudama Singh* in *Ajay Maken* (note 13 above) fortified it as a precedent.
health, right to education and right to food, including right to clean drinking water, sewerage facilities and transport facilities.’ Further, it held that slum-dwellers should not be seen as illegal occupants of land, whether public or private, but as ‘rights bearers whose full panoply of constitutional guarantees require recognition, protection and enforcement.’ Emphasising the need for due process, the court directed the state agencies to complete the survey and consult the dwellers. As far as possible in situ rehabilitation should be preferred. If that is not feasible, then whenever the agencies are in a position to rehabilitate the dwellers elsewhere, they should be given adequate time to make arrangements to move to the relocation site.181 These orders that called for engaging with the slum-dwellers before evicting them can be seen as an instance of judicial activism.

In India, the slum demolition and evictions are carried out under the Slum Areas (Improvement and Clearance) Act, 1956 which provides wide discretion to the government for declaring an area as ‘slum’ and ordering its clearance.182 Different state governments have made context-specific amendments to this law. For example, in Maharashtra, the Maharashtra Slum Area (Improvement and Removal and Redevelopment) (Amendment) Act, 1995 establishes the Slum Rehabilitation Authority (SRA). These authorities are bureaucratic in their structure and do not contain any mandate to conduct consultation or engagement with the occupiers and owners of these lands. Although the Development Control Regulations of Greater Mumbai 1997 require that 70 per cent of slum dwellers must consent to the slum rehabilitation scheme before the government declares the land as a ‘slum’ and puts it up for redevelopment, given the extent of housing indigency in Maharashtra, it may be argued that these authorities have not been sufficiently effective.183 Indeed, in a recent judgment184 the Supreme Court recognised the plight of 800 slum dwellers in Santa Cruz (Mumbai) who were incidentally the owners of the land (under a consent decree Wadia Trust, the original owner, had transferred the title to slum-dwellers’ cooperative society185) but had been embroiled in litigation among the rival builders and the SRA. Consequently, the slum dwellers had been denied permanent housing on their own land for over three decades. The court invoked its discretionary powers under Article 142 of the Constitution given the peculiarities of this case. It directed the SRA to ensure that fresh bids were invited and the project was completed in a time-bound manner. The SRA was required to evaluate the plans of every bidder and put the most beneficial plan in its opinion on voting before the slum-dwellers. Given the lack of agency to participate in the development of the plans, the slum-dwellers, although made part of the consultation, had limited say. The

181 Ajay Maken (note 13 above) 103. In Sudama Singh (note 180 above), the Delhi High Court made evictions conditional upon providing alternative accommodation. It borrowed insights from South African jurisprudence and held that a ‘meaningful consultation’ must be held with the slum-dwellers. Further, the Delhi Legal Service Authorities should organise periodic camps in slums to generate awareness about the rights of slum-dwellers.

182 Slum Areas Act ss 3, 4 & 9. Section 6 provides that the expenses for maintenance and repair of slums shall be borne by the occupiers who have little agency in deciding improvement or demolition.


184 Susme Builders Pvt. Ltd. v Chief Executive Officer, Slum Rehabilitation Authority, (2018) 2 SCC 230. The court notes: ‘There are many powerful persons involved, be they builders, promoters and even those slum dwellers who have managed to become office bearers of the society of slum dwellers.’ (Ibid at 239.)

185 The earlier Development Control Regulations of Greater Mumbai, 1991 required that 70 per cent of slum-dwellers should form a registered cooperative society before the Slum Areas Act could be invoked. These regulations were amended in 1997 to mandate consent from 70 per cent slum-dwellers instead.
court order meant that they would only be offered limited plans to choose from rather than ensuring their participation in the planning itself. This judgment lacks the crucial aspect of participation critical to the right to the city that the rule of ‘meaningful engagement’ seeks to do in the judgments of the South African Constitutional Court. The High Court of Delhi, nonetheless, has recently engaged with South African jurisprudence on the right to housing.

In *Sudama Singh*, people living in various slum clusters of Delhi for several years petitioned the High Court. They argued that demolition of their houses without relocation and rehabilitation was against the policy and the Constitution. The Government argued that the demolition was legal as these clusters were on the ‘right to way’. The court said that the slums were a result of an unbalanced growth where people in rural areas were forced to migrate to cities due to pressure on agricultural land and lack of employment opportunities. In cities, the formal accommodation is prohibitively expensive and formal employment is unavailable. The High Court drew from the judgments of the South African Constitutional Court in the *Grootboom*, *Olivia Road* and *Joe Slovo* cases to direct that the state agencies undertake a careful survey to identify the slum dwellers. The agencies should ‘meaningfully engage’ with those who are being evicted and provide basic civic amenities at the site of relocation. While the judgment is laudable, it fails to specify any structure for and terms of ‘meaningful engagement’.

Liebenberg’s apprehension in the South African context applies with even more force to the Indian context, given the general disregard of procedures in public interest litigation. Drawing from *Sudama Singh* and the South African jurisprudence, the Delhi High Court facilitated ‘meaningful engagement’ of various stakeholders and civil society members in *Ajay Maken*. The court noted that the right to housing is much wider than a right to basic shelter. It is against this backdrop that the court opined that there is a need to recognise the emerging concept of the right to the city. It noted that after the *Sudama Singh* judgment and the contempt petition following, the new 2015 policy and the 2016 protocol included the terms of the judgment and acknowledged that the right to housing includes the rights to life, education, healthcare, livelihood, food and water, clean environment and public transport. Thus, the constituent features of the right to the city are recognised by the 2015 policy, the court opined.

It must be acknowledged that both *Sudama Singh* and *Ajay Maken* are the exceptions rather than the rule. It highlights the worrying inconsistency among the Indian courts on the issue of housing rights. This impedes the development of any solid jurisprudence on housing rights. Very few judgments in India have actually engaged with historical and other factors behind such unequal cities. This also hides the complicity of the state and makes informal housing in slums and ghettos appear natural and chosen. The *Grootboom* and *Yolanda Daniels* judgments from the South African Constitutional courts provide good examples of using historical and sociological

---

186 *Sudama Singh* (note 180 above).
187 The petitioners contended that even those houses that were beyond the area marked for the widened road were demolished and, in most cases, people did not deliberately set up their shacks on the areas marked for roads. In fact, a lot of them had migrated to Delhi even before there were roads in that area. No earlier policy provided for any exceptions to rehabilitation such as the ‘right to way’.
189 While ‘meaningful engagement’ under the auspices of the High Court was successful in *Ajay Maken* (note 13 above) the lack of regulatory mechanism and allocation of funds, and an absence of any clear structure could derail such engagement. The 2015 policy refers to holding awareness camps and consultation with the slum-dwellers during the survey.
perspectives for a purposive interpretation of law. Such perspectives would help the courts to interpret the deeper purpose of these enactments not only from a point of view concerned with the welfare of individuals but importantly from a sense of corrective justice that the state owes to the marginalised races, castes, religious communities and other disadvantaged groups.

IV CONCLUSION

The limited sample of judgments reveals a tendency to give a broader interpretation to the right to housing. The courts of South Africa and India have both recognised the right to health, food and sanitation as aspects of the right to housing. In some cases, a nexus has been drawn between the right to livelihood and housing as well. This recognition is underpinned by the value of human dignity and equality that informs the constitutional interpretation in both India and South Africa.

It is noteworthy though that the suggestions of linkages with the right to the city are largely confined to demands of the pavement or slum dwellers to either stay in the city or be relocated and rehabilitated with adequate amenities and livelihood. The courts, as mentioned before, have held that housing rights are inter-related with other constitutional rights and the same extends to the dependent members as well. Predominantly, the concerns around housing have largely to do with the infrastructure and habitability aspects. Primarily, therefore, the emphasis has been on expansion of the instrumental elements of the right to the city that focuses on access to the facilities and amenities that a city offers.

On the transformational element of this right, the emphasis is on the democratic right of every citizen to participate and have a stake in the decisions of the city, and the right to appropriate the advantages of a city. The South African Constitutional Court has mandated that before conducting evictions, the state agencies should strive for ‘meaningful engagement’ with the potential evictees to ensure their participation in relocation and rehabilitation plans. However, ambiguity exists on the matter of its contours and details. The lack of regulatory safeguards around engagement is also a concern, and is perhaps the reason for an uneven application of the concept. In contrast, the Indian apex court has mandated neither rehabilitation nor any systematic due process and notice before eviction. It has only been discussed in a few judgments of the Delhi High Court and there too the issues of ambiguity and lack of regulatory safeguards persist. These sporadic judgments only highlight the inconsistencies in housing rights jurisprudence in India.

Notwithstanding certain advancements in the housing rights jurisprudence, we are far from anchoring the right to the city in constitutional law in a way that effectively prioritises the rights to appropriation and participation of people in shaping their city. The law has, at times, legitimised violence both in its enactment and enforcement. This has left the urban poor at the mercy of powerful interest groups who use legal instruments to categorise slums as ‘illegal’ and dispossess its dwellers of any right to residence. They are then forced to agree to
a ‘resettlement scheme’ in the fringes of the city. In India, this violence is enacted through statutes, executive orders and zoning regulations passed under the aegis of the central and the state governments. Further, the judicial insensitivity that labels slums as ‘encroachment’ or ‘pickpocketing’ of urban public lands has legitimised dispossession. The judgments are not informed by an analysis of the historical and sociological processes which, along with the complicity of the state, provide an explanation for the alarming levels of urban poverty and slums. An understanding based on the ‘right to the city’ would provide that nuance.

In India, the Supreme Court has never ruled that providing alternative housing and resettlement should be a condition precedent for any eviction plan. Thus, we see insensitivity and the violence of forced eviction legitimised through court orders on matters of ‘public interest’. The executive in some states, through their policies, have partly incorporated the conditional requirement of an eviction plan but even then, as seen in Ajay Maken, they permit arbitrary exceptions. At times, the courts have adopted language that prioritises property rights over the sustenance and livelihood of the slum-dwellers. The 138th Report of the Law Commission of India has emphasised the need for a statutory provision protecting the rights of the slum and pavement dwellers, particularly because the existing statutes give a wide discretion to the executive to identify and clear slums without an obligation to provide alternative accommodation. The Report noted that such a law should ensure that local authorities do not withhold essential services and civic amenities to the occupiers and should provide them with at least two months’ notice before eviction. The Commission recommended that central legislation be enacted to provide that ‘slum dwellers shall not be evicted without providing to them alternative accommodation’ and such accommodation should be within a short distance from where the slum-dwellers are to be evicted which allows them to access the city, as far as possible, in the same way as before. To this, one may also add the need for a structured and monitored engagement process to ensure participation of all the stakeholders to arrive at a fair settlement. However, given that there has not been any action on the recommendations of the Law Commission, it is difficult to be optimistic.

The South African Constitutional Court has been far more sensitive to the rights of occupiers, and the prospects for enshrining the right to the city in its jurisprudence on housing appears to be more positive. The de-emphasis of private property rights and entrenchment of the concept of ‘meaningful engagement’ does provide a voice for a poor occupier in the processes of (re)location. At the same time, the engagement with the social history behind the dispossession of land by black people and consequent insecure tenure, such as occurred

---


191 Almitra Patel (note 169 above).

in *Yolanda Daniels*, has provided the Court with justification for expanding the beneficial scope of s 25(6) of the Constitution, or ESTA not only as a part of corrective justice but also to enable people to renew their citizenship in the city. While *Yolanda Daniels* was only about a plea by an ESTA tenant to carry out certain essential improvements, the Court aided by a socio-historical perspective, could see a wider right-to-city implication to justify the improvement. Importantly, this also helps us subject the existing arrangement of property rights to critical inquiry.

The task of developing the content of the right to the city remains a challenge, and comparative studies can help us conceptualise this right in a manner that it not only remains a rallying slogan for influencing the politics of the state, but a legal tool to enforce important claims. The legal-normative work carried out in parts of Latin America and the World Charter on the Right to the City provide important insights for both the nations that can be supported by the value of human dignity in constitutional interpretation. Of course, the right to the city should at the least be consistent with the dynamism of the city itself. As Purcell argues, ‘the right to the city is not a panacea. It must be seen not as a completed solution to current problems, but as an opening to a new urban politics.’ This means that elements of the right may be enshrined in constitutional law, but it is also a framework that should not be rendered entirely rigid through legal doctrine but allow the space for politics as well.

Lastly, the efforts to make cities world-class, beautiful, and slum-free, the modern cities have turned out to be brutal exclusionary projects. They have not only prioritised the exchange value over the use value, but also limited the access to an urban life and its opportunities for impoverished people who literally fight for their location in the wider socio-economic structure. A principled recognition of the right to the city could ensure that these smart cities are also just, fair and sustainable.

---

193 *Yolanda Daniels* (note 144 above, at para 14). The Court, for example, notes, ‘Dispossession of land was central to colonialism and apartheid. It first took place through the barrel of the gun and “trickery”. This commenced as soon as white settlement began, with the Khoi and San people being the first victims. This was followed by “an array of laws” dating from the early days of colonisation. The most infamous is the Native Land Act (subsequently renamed the Black Land Act). Mr Sol Plaatje, one of the early, notable heroes in the struggle for freedom in South Africa who lived during the time this Act was passed, says of it, ‘Awaking on Friday morning June 20, 1913, the South African native found himself, not actually a slave, but a pariah in the land of his birth’… Apartheid sought to divest all African people of their South African citizenship.’

194 The Court observed: ‘If you deny an occupier the right to make improvements to the dwelling, you take away its habitability. And if you take away habitability, that may lead to her or his departure. That in turn may take away the very essence of an occupier’s way of life. Most aspects of people’s lives are often ordered around where they live. Bell says “[a] tenant who fears loss of an interest as vital as his home may forego associations or actions that are a normal part of self-determination and self-expression.’ (Ibid para 25).

195 Purcell (note 15 above) 99.
Dismantling Apartheid Geography: Transformation and the Limits of Law

RALPH MADLALATE

ABSTRACT: This article reflects on aspects of the Constitutional Court’s contribution to shaping spatial politics in post-apartheid South Africa. In a milieu characterised by extreme inequality and racial and class-based segregation, it interrogates the Court’s engagement with the odious legacy of apartheid geography which continues to lock South Africa’s cities and rural areas into its racist logic. The article begins by exploring the notion of apartheid geography as a product of racial, class-based and spatial discrimination buttressed by discriminatory laws. The article then outlines the relationship between law and space by drawing from critical legal geographic theory to cast South Africa’s space as the product of racial, class-based and spatial discrimination. As a key site in the legal production of space, this article analyses how the Court in Mazibuko v City of Johannesburg and Daniels v Scribante understands space in an effort to map its own role in reconfiguring post-apartheid spatial relations. This article introduces critical legal geographic methodology as a way of understanding the Court’s spatial jurisprudence in the post-apartheid era. Mapping the theoretical insights of this interdisciplinary approach to the study of law and geography reveals shortcomings in legal theory. This article argues that adopting critical legal geographic methodology will lead to increased recognition of the operation of spatial, racial and class-based inequities in legal adjudication. This may in turn lead to spatially contextualised judicial decision-making that aims to address apartheid geography.

KEYWORDS: race, cities, space, South Africa, apartheid geography, transformation, inequality

AUTHOR: Research Fellow, Harvard Law School, Centre on the Legal Profession Email: fmadlalate@gmail.com

ACKNOWLEDGEMENTS: I presented an earlier version of this paper at the Constitutional Court Review IX conference in Johannesburg, I wish to thank the participants for their comments. I am grateful to Vuyisile Ncube, Mpho Raboeane, Kelly Hwang and Jason Brickhill for useful comments on earlier drafts. I am also indebted to the editors and anonymous reviewers for their helpful suggestions.
I INTRODUCTION

Anyone who travels through our beautiful countryside cannot help but notice that the living conditions of workers who live on farms do not always meet a standard that accords with human dignity. There is little doubt that things have improved, but unfortunately not uniformly so. Why not? What does the promise of transformation mean for post-apartheid spatial relations? In the context of racial segregation and spatial inequality what role can the Constitutional Court (‘the Court’) play? This article reflects on aspects of the Court’s contribution to shaping spatial politics in post-apartheid South Africa. In particular, I interrogate the Court’s engagement with the odious legacy of apartheid geography which continues to lock South Africa’s cities and rural areas into its racist logic. This article develops in three main parts. Part I briefly explores the notion of apartheid geography, showing that it lies at the intersection of racial, class-based and spatial discrimination buttressed by discriminatory laws. I argue that the landscape of apartheid has socio-political effects which often go unacknowledged, not least, in the resolution of legal disputes. Part II outlines an approach to analysing the relationship between law and space by drawing from critical legal geography. By adopting a Lefebvrian view of space as a social product, the spaces we inhabit are brought to life and shown to be the subject of sustained conflict. In these sites of contest, law and legal processes play a dominant role which warrants interrogation. This part shows that critical legal geography provides a valuable theoretical frame which we can use to make sense of the Court’s spatial jurisprudence.

Part III then turns to consider some aspects of the Court’s spatial jurisprudence through the lens of critical legal geography. Juxtaposing the Court’s judgments in Mazibuko v City of Johannesburg, and Daniels v Scribante, this article introduces critical legal geographic methodology as a way of understanding the Court’s spatial jurisprudence in the post-apartheid era. I argue that by adopting this methodology we can more adequately recognise the operation of spatial, racial and class-based inequities in legal adjudication. This will in turn lead to spatially contextualised judicial decision-making in response to apartheid geography. This part revisits Mazibuko and recasts it as more than a purely socio-economic rights case, although it has been rightly received as a key case in the socio-economic rights canon, I argue that the case is equally important for what it demonstrates about the Court’s understanding of post-apartheid space. I then consider the Court’s understanding of space in Daniels, arguing that this decision edges towards a critical understanding of apartheid geography that differs from the Court’s approach in Mazibuko. As a key site in the legal production of space, I assess how the Court in these two judgments understands space in an effort to map its own role in reconfiguring post-apartheid spatial relations. To be sure, the focus on two judgments does not decode the Court’s entire record, and is not conducive to drawing definitive inferences. I focus here on tentatively discerning operative themes in the judgments. In so doing, I explore the workings of space, race and law in post-apartheid South Africa; and illustrate the utility of critical legal geographic approaches in this context. Undoubtedly, dismantling the numerous manifestations of apartheid geography cannot be achieved by the mere rendering of legal decisions. However, I focus here on decisions of the apex Court in order to understand some of its normative interventions in the light of the country’s racialised spatial inequality. A complete

1 Daniels v Scribante & Another [2017] ZACC 13, 2017 (4) SA 341 (CC)(‘Daniels’) para 111.
2 Mazibuko & Others v City of Johannesburg & Others [2009] ZACC 28, 2010 (4) SA 1 (CC)(‘Mazibuko’).
3 Daniels (note 1 above).
examination of the Court’s interventions is beyond the scope of this article although certainly worthy of further study.

A Decoding apartheid geography

What I refer to as ‘apartheid geography’ has origins that precede the rise of formal apartheid. Indeed, the existence of racially distinct spaces can be traced to the colonial encounter. By apartheid geography, I mean the creation and maintenance of racially-identified spaces, coupled with racial and class-based segregation and an uneven distribution of social goods and public amenities, which are skewed in favour of white people. In essence this phenomenon is an expression of white supremacy, concretised and given meaning through laws and policies which produce physical, psychological and social barriers. This process is premised on segregating space to determine access to social goods and opportunities. This logic was central to the creation of spaces to privilege white users in apartheid South Africa.

The production of apartheid geography is characterised by the vicious manipulation of black people’s access to space through racist social and economic policy crystallised in an array of discriminatory laws. Indeed, much of the story of black space is characterized by marginalisation, involuntary migration, insecurity of tenure and legal precariousness. Much of the legal basis which has produced this phenomenon predates formal apartheid yet adopts a near identical logic which produced harm in equally violent ways. In the urban context, this logic is manifest in cities characterised by a commercial or industrial centre occupied by predominantly white-owned businesses and enterprises which are skirted by poorly-serviced and overcrowded township areas designed for occupation by an overwhelmingly black labour force. In the countryside, this logic is apparent in the large numbers of black people who occupy spaces in which they have insecure tenure as former sharecroppers turned labour tenants, farm dwellers and occupiers of land in which they hold limited rights.

While racial segregation was a feature of colonial life in the nineteenth century, this project gained momentum in the twentieth century. This period saw black peoples’ rights in land eroded and extinguished as the state led by a white minority created a country in which black people had severely limited rights. Early changes in the South African economy were facilitated by legislation such as the Glen Grey Act 25 of 1894 which created distinct areas for black people while simultaneously directing their economic lives. The thrust of this law is encapsulated by Cecil John Rhodes who, speaking at the Cape House of Parliament, asserted:

Every black man cannot have three acres and a cow, or four morgen and a commonage right. We have to face the question, and it must be brought home to them that in the future nine-tenths of

---

6 I use ‘black’ to refer to Indian, Coloured and African people.
10 N Worden (note 4 above) at 54.
them will have to spend their lives in daily labour, in physical work, in manual labour. This must be brought home to them sooner or later.\textsuperscript{11}

At a spatial level, this exploitative project gained momentum with the enactment of the Natives Land Act 27 of 1913 and the later Native Trust and Land Act 18 of 1936, which barred black people from owning 87 per cent of the country’s land. This segregationist urge was coupled with a desire to extract from black people their labour and partly incorporate them into the country’s economy as sources of cheap labour.\textsuperscript{12} Infamously, these laws and other measures facilitated the dispossession of black people of their land and its reallocation to white farmers and settlers. These laws also confined the African population to woeful conditions in rural reserves which were politically and socially isolated from parts of the country designated for white people. These laws tore black men from their livelihoods as subsistence farmers and thrust them into wage labour either on farms or in mines, while black women were condemned to languish in the rural reserves trying to eke out a living on the meagre resources that were available. These rural reserves became sites of concentrated poverty, unemployment and malnutrition.\textsuperscript{13}

In urban areas, the Natives (Urban Areas) Act 21 of 1923 ensured that racial integration would be strictly controlled and that the settlement patterns of black people would take place on terms dictated by the white government. The rationale of this scheme is captured in the report of the Stallard Commission which looked into racial integration in urban areas and concluded that:

The native should only be allowed to enter the urban areas, which are essentially the white mans’ creation, when he is willing to enter and to minister to the needs of the white and should depart therefrom when he ceases to so minister.\textsuperscript{14}

Thus the Stallard doctrine laid the basis of racialised spatial inequality in South Africa’s growing cities.\textsuperscript{15} It meant that black people were constantly haunted by the spectre of unlawfulness when they entered white zones, including urban areas and some rural areas from which they could be ejected with few procedural and substantive safeguards. Within this racist logic, the pattern of urbanisation in the twentieth century was set. Black people’s spatial practices were valid only to the extent that they served the interests of white South Africa. Traditional ways of being were violently disrupted and replaced by a state-led spatial practice which reinforced white supremacy.

During this period, black spatial practice (including traditional livelihoods, customary practices and secure tenure) was put under mutually reinforcing pressures. The loss of land rights in rural areas coupled with the implementation of taxes which forced black people into wage labour resulted in migration into urban areas where black people had limited temporary rights and could thus be exploited as mine workers or labourers in the country’s emerging


\textsuperscript{12} H Wolpe (note 9 above) at 440.


\textsuperscript{14} R Davenport ‘African Townsmen? South African Natives (Urban Areas) Legislation through the Years’ (1969) 68 African Affairs 95, 95.

\textsuperscript{15} This follows a pattern of racist urbanisation in colonial societies (W Shaw \textit{Cities of Whiteness} (2007) 37).
industries. This placed black spatial practice under strict control, as Beinart and Delius rightly argue:

The Land Act prohibited purchasing of white land, but not occupation; it demarcated scheduled areas for African communal land ownership; and it controlled the forms of tenancy on white-owned land. Agricultural production on white land was controlled, not prohibited. While each region revealed its own patterns, almost all had one thing in common: white-owned land was not occupied exclusively by whites. Black people predominated on the great majority of white-owned farms in South Africa, where they lived as tenants and workers both before and after the 1913 Land Act. These spatial effects were undergirded by a racist legal environment. The change in patterns of land ownership, coupled with the fact that black people continued to reside on land given over to white owners helped create a discourse of spatial illegality which would become the hallmark of apartheid. In this process, the law was used to determine who could be present, as well as when and on what terms this was to occur. By using the law in this way, legal concepts such as ownership, nuisance, trespass and eviction became tools wielded by the state and private actors to advance a racist agenda and ensure that the formation of South Africa’s urban and rural areas occurred on terms beneficial to white people. In this way, the law both sowed the seeds of, and upheld, racialised unlawful occupation in the country.

The processes and mechanisms determining lawful access to space based on race reached increased intensity and scope with the rise of the apartheid government. This period saw the furtherance of the state’s racist policy through the enactment of laws which would be zealously advanced by the National Party through the Group Areas Acts of 1950 and 1966 which ensured that urbanisation was within the racial classification paradigm established by the Population Registration Act 30 of 1950. This marked an important period in the development of the country’s spaces. The implementation of the Group Areas Act meant that any semblance of racial integration within cities could be curtailed and black residents forcibly removed to townships designated for them. These forced removals were coupled with influx control measures designed to regulate the racial demographics of space at any given time. Effectively, the state could determine the number of black people required as labourers in an urban area and expel those deemed surplus to requirements. This ruthless system was legally enforced with the courts playing a key role in its implementation. The law's harsh operation in the removal of black persons from white areas is captured by Didcott J:

Once you are officially ‘idle’, all sorts of things can be done to you. Your removal to a host of places, and your detention in a variety of institutions can be ordered. You can be banned forever from returning to the area where you were found, or from going anywhere else for that matter, although you may have lived there all your life. Whatever right to remain outside a special ‘Bantu’ area you gained by birth, lawful residence or erstwhile employment is automatically lost.

The policing and manipulation of space in line with the state’s policy occurred on at least three mutually reinforcing scales. First, at the nano-level: as was evident in restrictions on interpersonal relations. Secondly, at the micro-level, as was evident in the implementation of urban group areas and household segregation. Thirdly, at the macro-level, as was evident in the interdependent yet unequal relationship between the Bantustans, the reserves and white South

18 E Unterhalter (note 13 above) at 12.
19 In re Dube 1979 (3) SA 820 (N) 821C–821E.
Africa’s urban centres. This system of spatial control made black people legally vulnerable in cities, major industrial and commercial centres in relation to white South Africans. They would eventually be denied political membership and citizenship from these spaces through the Bantu Homelands Citizenship Act of 1970. The operation of these laws meant that black people entered South Africa from the Bantustans as temporary sojourners whose notional political citizenship lay in the ‘independent’ Bantustans.

While this vignette does little to capture the callous indignity of black life within apartheid geography, it illustrates that spatial expressions of racial discrimination formed a key pillar of the state’s racist policy. These laws created areas which differentiated their users’ experiences on the basis of race and ensured that inequality would be reproduced by their very design. This is evident in the vast distances between areas designated for black occupation and central business districts, public schools and health-care facilities that were entrenched by apartheid spatial planning. Contemporary manifestations of this spatial discrimination are evident in the large numbers of black people experiencing precarious tenure in land and housing. This spatial dislocation also has economic consequences as the occupants of these racialised spaces are simultaneously dislocated from education and work opportunities, which are key to upward mobility. Accordingly, apartheid geography is far more than an historical aberration or a mere infrastructural problem – although this is evident in the country’s chronic housing shortage. Instead, it is a social condition premised on the reproduction of a racial hierarchy through spatial means. Clearly the law, through a series of directives, interdictions and bans, has been central to this process. This makes the geography of apartheid at once a geography of law: a network of legal provisions which shape use and access to space, thus guiding socio-political outcomes.

To be sure, discriminatory laws have been central to the production of South African space from the colonial encounter, to the zenith of apartheid. While these laws have been repealed and replaced with new legislation in line with the country’s non-racial ‘transformative’ constitutional project, what they have left behind are spaces indelibly marked by the politics of racial oppression. This tenacious expression of racism has proved difficult to dislodge in the post-apartheid era. This is despite a raft of new laws promulgated in response to the history of black spatial insecurity. While these pieces of legislation are designed to enhance security of tenure and move the country away from apartheid geography they are frequently pitted against private interests in property. In this milieu, the legal concepts of nuisance, trespass and eviction continue to play a prominent role in regulating black spatial practice. The courts have been called on to resolve the tension between old order rights and socio-economic rights.

---

21 E Unterhalter (note 13 above) at 19.
22 Statistics from June 2018 show that 13.6 per cent of South Africans live in informal dwellings, and 5.5 per cent in traditional dwellings. Stats SA. The latest household statistics and more available at http://www.statssa.gov.za/?p=11241.
guaranteed to unlawful occupiers.26 This has generated a vibrant discourse centred on the intertwined rights to property, housing and related rights in South African legal circles.27 Much of this literature is framed using the language of rights. Recent scholarship has sought to incorporate Lefebvrian ideas around the ‘right to the city’ into notions of rights-based spatial practice.28 Responding to the lack of race consciousness in this literature, I have argued elsewhere that legal discourse is inadequately attuned to the inequity of racialised space in resolving contestations over rights in urban settings.29 In what follows, I introduce critical legal geography as a methodological approach to unpacking the complexity of the intersection of race, geography and law in post-apartheid South Africa. This approach, I argue, has the potential to improve legal responses to the effects of apartheid geography.

II LAW’S SPACE: UNDERSTANDING LEGAL GEOGRAPHY

This part draws on the growing literature concerned with legal geography, and mapping the theoretical insights of this interdisciplinary approach to the study of law and geography. This is in order to capture the contours of the relationship between space and law; and to understand its importance to human existence. This reveals the social and political meaning of law and its operation in the context of space.

A The politics of space

Critical legal geography is premised on a contextual approach to space. In what is a foundational contribution to the field, Henri Lefebvre pioneered an approach to conceptualising space which escapes the commonplace view of space as a mere natural container in which life takes place.30 On the contrary, Lefebvre argues that it is the very processes of life which serve to produce space. Accordingly, space is not merely inert matter but a living social construct – deliberately created pursuant to social action. In this view, ‘the spatial practice of a society secretes that society’s space’ in a gradual process.31 Thus, Lefebvre asserts that space and time are inextricably linked to social and political practice:

Space is not a scientific object removed from ideology or politics; it has always been political and strategic. If space has an air of neutrality and indifference with regard to its contents and thus seems to be ‘purely’ formal, the epitome of rational abstraction, it is precisely because it has already been occupied and used, and has already been the focus of past processes ... Space has been shaped

31 Ibid.
and moulded from historical and natural elements, but this has been a political process. Space is political and ideological. It is a product literally filled with ideologies. According to Lefebvre, spatial conventions far from being neutral tend to reveal the social relations which produce them while creating or transforming existing social relations. Demystifying socio-spatial relations allows us to decode these processes and reveals the forces that underpin them. This calls us to resist thinking of space as pre-ordained, natural or ahistorical. Instead, we see social and spatial relations as inextricably linked in an interdependent relationship which is both ‘space-forming and space-contingent’. In this view, society shapes space and is simultaneously shaped by space. This occurs as space becomes an expression of social policies, prejudices and ideals. Viewed through this lens, phenomena such as mass evictions, gentrification and the displacement of racial and ethnic groups, are not isolated events but part of a socially driven narrative which Lefebvre termed ‘the production of space’. In my view, the contemporary production of space is channelled through an array of laws and legal norms which direct spatial practice. These include housing and eviction laws, zoning and spatial planning regulations and trespass laws. These laws and legal norms often mask a variety of other interests, such as maintaining racial and class-based divisions, or excluding segments of the population, deemed otherwise undesirable, from certain spaces.

This socio-spatial theory articulates an intuition understood by the architects of racial segregation and apartheid in South Africa. Apartheid geography operates by denying access to what Galster terms ‘opportunity structures’. By limiting access to opportunity structures, the apartheid state was able to determine black people’s lives by confining them to underserviced, overcrowded areas with poor access to health-care and education facilities. At its core, this project harnessed the power of space to define its users’ lives in key political, economic and social ways. This process, in turn, reinforced racist ideology. By shaping the content of black life, space was central to producing racial identities in South Africa. Blackness was forged within sites of concentrated poverty and social isolation which restricted individuals to few opportunities save to be exploited as cheap labour. In contrast, whiteness was shaped by access to capital, both economic and social, health, wellness and technology which allowed its beneficiaries widespread opportunities for upward mobility. This racial dichotomy was incubated within South Africa’s segregated spaces. As an instrument of social control, apartheid geography served its purpose and continues to differentiate its users’ experiences in ways that have yet to be undone.

Understanding space as a social product calls us to resist relying solely on legal categories. Instead, it demands decoding spatial politics beyond legal doctrine. This reveals that space operates in ways that are irreducible to legal categories. For instance, land, housing and

---

33 H Lefebvre (note 30 above) at 81.
34 Ibid.
environmental rights are often viewed as distinct (legal) issues in the South African context regulated by separate pieces of legislation and policies documents. This obscures the fact that natural, manufactured and social spaces are intertwined. I thus consider the different zones – urban, rural, public and private – under the banner of space which captures the range of considerations relevant to each without viewing them as necessarily disparate. I adopt this approach, as the spatial politics of the rural and urban are inextricably linked in the South African context such that we cannot properly understand one without understanding the other. This is due to the interdependent relationships between homelands and cities, black townships and white industrial and commercial districts. Accordingly, contemporary South African public discourse, which is increasingly dominated by contestations over land, cannot be separated from questions of racial justice in housing and access to resources in cities.

B Mapping the law: introducing critical legal geography

The influence of Lefebvrian constructions of space encourages understanding law and geography as co-constitutive. Viewing space as a social product illustrates the manner in which space, law and society are related. This reveals that ‘law and space actively shape and constitute society, while being themselves continually socially produced’.\(^{39}\) This co-constitutive view of law, society and space distinguishes critical legal geography as an analytical approach. Once recognised, this interaction challenges us to rethink both law and space as Blomley et al assert, ‘by reading the legal in terms of the spatial and the spatial in terms of the legal, our understanding of both “space” and “law” may be changed’.\(^{40}\)

Once we view space as a social construct produced through everyday interactions, we can critically examine the landscapes of legal geography. This school of thought draws on critical theory, viewing law as a discourse of power which is constructed and contested within society.\(^{41}\) In line with this approach, our analysis is not ‘confined to a doctrinal analysis of legal reasoning set out in the text of the case, law report or judgment’.\(^{42}\) Instead, the analysis goes beyond the text to locate law in its social context – mapping law onto its place in everyday life. This means expanding the notion of what is relevant in legal analysis to include questions about the spatial location of litigants, the milieu they occupy and other markers, be they political, economic or social, which constitute their identities and determine their social positioning. This wider view seeks to capture a broader range of situational influences operative in the resolution of legal disputes. This approach resonates with the ‘call to context’ issued by John Calmore, who argues that: ‘traditional legal analysis and advocacy are too often plagued by the tendency to extrapolate issues from their history and the broader social and normative contexts that bear so heavily on them’.\(^{43}\) This view recognises that legal disputes are rarely ‘just legal problems’ but arise in a context marked by race, class and space.

At the core of the critical legal geographic approach is an attempt to denaturalise the operation of legal arrangements which have gained the appearance of being irrefutable truths. This reveals that seemingly neutral legal moves are underpinned by particular spatial

\(^{39}\) S Blandy & D Sibley ‘Law, Boundaries and the Production of space’ (2010) 19 Social & Legal Studies 275, 278.


imaginaries: both articulated and unspoken spatial visions motivating actions of interpretation and delimitation. Understanding this necessitates viewing law as nested within a range of relationships in which the role of law is but one strand in a complex web of elements. Thus, legal geographers engage with different ‘ways of knowing the world’ in order to capture the range of experiences which arise at the interface of law and society. This approach resists the reification of law and reveals that beneath law’s magisterial appearance lie numerous sites of contestation. These contestations are often shrouded in silence as positive legal texts become the dominant expressions of socio-legal reality. In order to escape this shortcoming, legal geography goes beyond legal texts, exploring the telling silences about space in law. Thus, rather than beginning with the text, critical legal geography calls us to locate the law.

Scholarship in critical legal geography observes that while law is everywhere, law’s operation is not always cognisant of where it operates. This lacuna is explored by several scholars. For example, Blomley and Bakan argue that:

> The legal mentality is curiously acontextual, such that legal relations and obligations are frequently thought of by the Courts and other legal agencies as existing in a purely conceptual space, with little recognition of their spatial heterogeneity or the local material contexts within which law is understood and contested.

Similarly, in the context of political boundaries, Richard Ford argues:

> There is no self-conscious legal conception of political space. Most legal and political theory focuses almost exclusively on the relationship between individuals and the state. Judges, policymakers, and scholars analogize decentralized governments and associations either to individuals when considered vis-à-vis centralized government, or to the state, when considered vis-à-vis their members, but consider the development, population and demarcation of space to be irrelevant. Space is implicitly understood to be the inert context in which, or the deadened material over which, legal disputes take place.

Following this line of critique Bartel et al argue: ‘that law does not transcend place, but is still dependent on it, is not a truth generally acknowledged by law’s servants and scholars’. As a result, legal concepts are often inadequately spatialised. The preference amongst lawyers is towards reliance on abstract principles which exclude from legal consideration issues such as location, the effects of conquest, structural discrimination, or biased economic systems. Of course, competing views are never completely banished as dissenting narratives survive their formal exclusion, existing in oppositional discourse. Consider the contemporary call for ‘land reform from below’ and other subversive practices by marginalised groups such as Abahlali baseMjondolo. Given the powerful impact of spatial engineering in the country’s history, the tendency towards legal dislocation (the extrication of space from law) presents a serious challenge. This is so not least because the constitutional matrix creates a rights

---

44 L Bennet & A Layard (note 42 above) at 412.
46 L Bennet & A Layard (note 42 above) at 413.
framework, access to which depends largely on a person’s connections to space. Indeed, rights-claimants often experience the deprivation of rights based on where they are located. For example, urban residents of black townships and informal settlements demand housing, water and other basic socio-economic rights while the deprivation of those resident in South Africa’s rural areas is compounded by their geographic remoteness from basic education, health-care and social security infrastructure.

C Towards a synthesis of space, society and law

So far, I have discussed space, law and society as distinct, and in so doing, I have been echoing the dominant divisions of labour which shape discourse on these issues. This is, of course, not an accurate reflection of everyday life. The work of critical legal geography is to weave together the seemingly separate discourses to make sense of their interconnectedness. In adopting this approach, I follow Blomley who argues: ‘law and geography do not name discrete factors that shape some third pre-legal, aspatial entity called society. Rather the legal and the spatial are, in significant ways, aspects of each other’. This perspective calls for integrated conceptions capturing the many complexities of social and legal and spatial relations. Developing a language to capture the points at which law, society and space intersect, Blomley deployed the term ‘splice’. For Blomley, this socio-legal concept identifies the ‘instances or moments where legally informed decisions and actions take place [in the sense of both of the occurrence of a legal performative (an event) and of being spatially located and embodied]. Splices are locally enacted encodings, which weave together spatial and legal meanings.’ Put differently, a splice refers to the notion of a simultaneous convergence of legal, social and spatial categories.

In everyday life, splices operate as powerful socio-legal frames which become naturalised and considered expressions of some common-sense idea. Blomley illustrates the legal and geographical interconnectedness of splices arguing that while some orderings are apparently based on the discourse of law, such as ‘citizen’, ‘these are simultaneously bound to spatial frames in ways that are mutually dependent’. These splices reflect the convergence of legal and spatial orderings which have been naturalised such that they lose their links to the discourse of law and space and appear instead as inert and pre-existing. For instance, ‘a legal category such as “citizen” is meaningless without the spatial category “territory”; similarly, the term “refugee” is considered as a legal categorisation but fundamentally reflects spatial dislocation’. The tendency to dislocate and naturalise these frames has resulted in them being understood as neutral legal categories apart from their spatial dimensions. This naturalisation has important consequences as Blomley argues:

54 L. Bennet & A Layard (note 42 above) at 410.
55 Ibid.
56 N Blomley (note 53 above) at 30.
57 Ibid.
A splice can appear simply part of the order of things, and thus non-negotiable. In so doing, splices can have a number of effects. Put bluntly, they construct the world in ways that systemically favour the powerful: employers, men, whites, property owners.\textsuperscript{58}

Thus, these splices encourage particular ways of thinking which legitimate spatial distinctions in governance and regulation. By marking a space as distinct, it can then be afforded different treatment not permitted in another space. In this way, people at the margins of society endure violence which has been naturalised and expected. Consider, for example, a commonplace splice in post-apartheid South Africa – the unlawful occupier. This frame is itself a new spatio-legal concept developed by the Court in interpreting the Prevention of Illegal Eviction Act 19 of 1998. In \textit{Port Elizabeth Municipality} the Court traced the terminological shift in the legislation from ‘squatters’ to ‘unlawful occupiers’ and the ideological shift from preventing squatting to preventing unlawful eviction which followed the repeal of the Prevention of Illegal Squatting Act No 52 of 1951.\textsuperscript{59} The swift integration of the term in the country’s spatial lexicon illustrates the power of law in generating legally reinforced social narratives. This conceptual reframing has not, however, fundamentally altered the nature of the condition described or freed it from its negative connotations. The concept of an unlawful occupier simultaneously refers to a legal status, social positioning and a spatial displacement which combine to produce results not attributable to either of these factors independently. Legally, unlawful occupation entails limited rights in property and implies a range of consequences. Spatially, it entails residence in a place which is not one’s own – a territorial dislocation. Socially, this carries a stigma which when combined with racial and class connotations in the South African context create a powerful social ordering frame. This framing reflects a spatial and legal precariousness, which, once naturalised, legitimates the legal and ethical distinctions made against unlawful occupiers and provides ‘justification’ for their (mis)treatment.

As I have illustrated, contemporary patterns of mass unlawful occupation exist as the product of colonial and apartheid racialisation, segregation and land deprivation. The concept of unlawful occupation can be traced to the shifts in spatial practice towards white individual ownership and the disruption of commonage-based spatial practice led by the state in the nineteenth and twentieth centuries. By removing many black people’s legal basis for occupation and confining them to specific places, the state created a spatial and legal category which allowed for expedited forced relocation and racial spatial engineering. Racialised shifts in land holding have made vast numbers of black people vulnerable to unlawful occupation as their land rights were extinguished while they continued to reside on the land of the new white owners. These unlawful occupiers could then be swiftly removed pursuant to common law principles.\textsuperscript{60} Considering the still unfolding effects of these processes, the idea of the ‘unlawful occupier’ as a neutral signifier becomes unsustainable. While anyone can theoretically bear this

\textsuperscript{58} Ibid.

\textsuperscript{59} \textit{Port Elizabeth Municipality v Various Occupiers} [2004] ZACC 7, 2005 (1) SA 217 (CC) (‘\textit{Port Elizabeth Municipality}’) at paras 11–13. The Court in \textit{Occupiers of Mooiplaats v Golden Thread Ltd} [2011] ZACC 35, 2012 (2) SA 337 (CC) at para 4 underscored the rationale behind this normative shift. Reprimanding Golden Thread for citing the occupiers as ‘the people who intend invading the Farm Mooiplaats’ and ‘the people who invaded the Farm Mooiplaats’, the Court held: ‘this description of human beings is less than satisfactory and cannot pass without comment. It detracts from the humanity of the occupiers, is emotive and judgmental and comes close to criminalising the occupiers. This form of citation should not be resorted to. A more neutral appellation like “occupiers” might well be more appropriate’.

\textsuperscript{60} \textit{Graham v Ridley} 1931 TPD 476.
description, the reality is that unlawful occupation continues to track racial, class-based and spatial distinctions drawn in the making of the country’s contemporary spatial arrangement. The notion of the unlawful occupier in the post-apartheid era is the conceptual successor of the racialised transgressor of spatial laws during apartheid. These persons would have been pejoratively labelled squatters under apartheid; they comprise people characterised as idle or redundant Africans – who were banished to the reserves – or urban blacks – who provided cheap labour in the urban economy. Post-apartheid jurisprudence reveals that many of the trajectories of these ‘unlawful occupiers’ have remained unaltered.61 This is true of many low-income residents of Johannesburg’s inner city.62

Once characterised as unlawful occupiers, the socio-spatial dimensions of this predicament are replaced by a convenient legal frame which obscures the subjects’ origins, history and lived experience. This framing simplifies the resolution of spatio-legal problems by circumscribing the range of considerations which are deemed relevant in adjudication which in turn leaves these splices uninterrogated. By adopting a critical legal geographic lens, we can challenge this dominant understanding pervasive in spatial jurisprudence and prompt increased attention to South Africa’s production of post-apartheid space. This methodology goes beyond traditional legal reasoning and relies on more than is contained in the parameters of legal disputes as characterised by the parties or adjudicators. By weaving together the various social, legal and spatial complexities which occur simultaneously and go unnoticed in everyday life, we arrive at an understanding which is concerned with far more than the rights a person enjoys in property, for instance, but a picture of reality which reflects the subjects’ true conditions. This view can potentially denaturalise oppressive social orderings that have gained the semblance of neutrality. In adjudication, this methodology can potentially improve the way that cases are adjudicated by widening the range of considerations which guide decision-makers.

In the preceding sections, I have provided a synopsis of apartheid geography showing that it is constantly reproducing hierarchies as society operates within preformed spaces while modelling future developments based on this racial logic. I then outlined a way of decoding the legalities surrounding this phenomenon in ways that go beyond static readings of space as pre-ordained. In this way I have shown some of the politics of interpretation, argument and framing which are central to how we make sense of our environments. In what follows I develop this idea by exploring the manifestations of apartheid geography in legal discourse.

III THE CONSTITUTIONAL COURT AND ITS RECOGNITION OF SPACE

While there are a range of registers available for socio-spatial contestations in the polis, the discourse of law occupies a dominant role.63 In this milieu, courts make authoritative interpretations of the meaning of fundamental rights. Naturally, the interpretations given to rights by the judiciary play an important role in the politics of space. The Court, as final decision-maker on fundamental rights in the South African Constitution, is thus a key site in the process of interpretation, representation and narration that shapes spatial discourse. Unsurprisingly, apartheid geography is a recurring theme in many cases before the Court,


63 South Africa’s constitutional democracy affirms the Constitution as the supreme law. (Constitution, s 1(c).)
albeit often as an unacknowledged backdrop. It forms part of the considerations in areas such as housing,\textsuperscript{64} education,\textsuperscript{65} and local government law.\textsuperscript{66}

In this part, I analyse contestations over legal meaning and contestations around space, tracking how post-apartheid space is to be understood and what spatial imaginaries are envisaged for the future. I juxtapose Mazibuko and Daniels, contrasting the approach to space contained in these decisions.\textsuperscript{67} I read these judgments through a critical legal geographical lens and suggest that this reveals far more than the meaning of the law. In so doing, I eschew an approach based on technical law, instead viewing law as a social construct in order to map the boundaries which the Court sets in adjudication at the intersection of race and space. By recognising that law is a contested system of knowledge which is a site of power struggles we can explore which discourses are most powerful in each case. While these decisions primarily involve the interpretation of socio-economic rights, I recast them as part of the legal production of space.

A Mazibuko and the legal discourse of space

Critical legal geography is a useful tool to weave together the composite experiences which occur at the intersection of law, society and space. Viewing legal cases through this lens reveals insights not ordinarily apparent in conventional legal theory. Adopting this methodology, I propose another way of reading one of the Court’s seminal space-related judgments. In Mazibuko v City of Johannesburg, the Court was confronted with the country’s legacy of racial segregation and the associated underdevelopment of black spaces. By all accounts, this case is concerned with the right to water; however, adopting critical legal geographic methods we can locate this judgment in its socio-spatial context revealing the racial, legal and geographic assumptions which underpin it.

In Mazibuko, several residents of Phiri, a black township in Soweto, brought a claim challenging the City of Johannesburg’s Free Basic Water policy on the basis that it failed to meet the constitutional guarantee of a right to have access to sufficient water. The challenge was premised on the argument that the impugned policy failed to provide the applicants with sufficient water given the specific needs of the affected community. In particular, the applicants pointed to the reality of overcrowding, multifamily residences and backyard dwellings in spaces originally designated for single-family use.\textsuperscript{68} These are contemporary manifestations of apartheid geography which, through influx control and other measures, confined black people to limited spaces. These spaces were underserved yet highly prized due to their proximity to the meagre economic opportunities available.\textsuperscript{69} The applicants further alleged that the limits placed

\textsuperscript{64} Government of the Republic of South Africa & Others v Grootboom & Others [2000] ZACC 19, 2001 (1) SA 46 CC; Port Elizabeth Municipality (note 59 above).

\textsuperscript{65} Federation of Governing Bodies for South African Schools (FEDSAS) v Member of the Executive Council for Education, Gauteng & Another [2016] ZACC 14, 2016 (4) SA 546 (CC) at para 38.


\textsuperscript{67} A full exploration of the Court’s spatial jurisprudence is beyond the scope of this article, though certainly worthy of analysis in further work.

\textsuperscript{68} Mazibuko (note 2 above) at para 88.

\textsuperscript{69} K Beavon (note 62 above) 231.
on their water usage were racially discriminatory in comparison to white neighbourhoods where the meters were not installed. This case illustrates the contemporary effects of not only the country’s racially tainted capitalism, which condemns many black people to structural unemployment, but also apartheid geography which ensured that access to housing and water in black spaces was woefully inadequate.

In resolving the dispute, the Court invoked a particularly de-contextualised reasonableness analysis finding that the policy fell within the bounds of reasonableness and, as such, was compliant with the right to sufficient water. Of interest in this article, however, is the manner in which the Court depicts space in the context of apartheid geography which continues to stratify our cities. In representing apartheid geography the Court adopts a historical framing strategy throughout the judgment. This framing stops short of interpreting the contemporary ramifications of this history. While exploring the influence of apartheid on the applicants’ claims, the Court frames the spatial context as follows:

Apartheid urban planning did not permit black people to live in the same urban areas as white people. Soweto was developed in accordance with this appalling racist policy. It is home to approximately a million people. Phiri, where the applicants live, is one of the oldest areas in Soweto. Most of the houses in Phiri are brick yet generally the people who live in Phiri are poor.

Evidently, the Court recognises the racially discriminatory motive behind the production of Soweto, and Phiri in particular. However, this framing is couched in historical terms, adopting a temporal lens rather than a spatial one. Emphasising the official fall of apartheid, the Court paints apartheid geography as a relic of the past. This obscures the social production of space which is shaped by repealed discriminatory laws such that, although formal apartheid has ended, the deleterious effects of racialised spaces are reproduced in contemporary spatial relations. Despite black spaces remaining disadvantaged by and large, the Court is keen to confine the effects of apartheid geography to a particular time without drawing connections between the contemporary vulnerability of black people and its historical origins in colonial and apartheid oppression. This framing effectively circumscribes the consideration of modern apartheid geography and allows the Court to foreground other considerations such as the need to allow the City to implement its policy. By describing history without drawing connections to the contemporary geographical manifestations of this history, the Court occludes a key factor in the applicants’ disadvantage.

The adverse effects of the historical framing of spatial disadvantage are also evident in the Court’s approach to the question of discrimination. The Court held that there was no racial discrimination on the part of the municipality by choosing to impose pre-paid water meters only on the residents of Soweto and not imposing similar measures on the city’s white suburbs or other ‘areas with poor black residents’. The Court thus dismisses the claim of discrimination arguing ‘it is not clear that the applicants have established that the policy impacted more adversely on black and poor customers given that other deemed consumption areas where poor black customers reside were not targeted’. The Court’s framing of the inquiry into discrimination also reveals some of the rationale behind the decision:

71 Mazibuko (note 2 above) at para 10.
72 Ibid at para 149.
73 Ibid.
To determine whether the discrimination was unfair it is necessary to look at the group affected, the purpose of the law and the interests affected. In this case, the group affected are people living in Soweto who have been the target of severe unfair discrimination in the past.\textsuperscript{74}

Clearly, the Court constructs racialised spatial disadvantage as a relic of the past. This narrative casts space as pre-ordained and does little to re-imagine an urban environment without apartheid geography. Instead, the Court accepts this historicised representation of apartheid geography and reproduces it. The Court reached this finding despite the City of Johannesburg conceding that the manifestations of racialised space were such that the geographic differences could result in racially disparate outcomes.\textsuperscript{75} By locating apartheid discrimination as a historical aberration confined to a particular point in time, the Court is able to characterise the applicants as the victims of past unfair discrimination. This interpretation effectively confines apartheid geography’s racial effects to the historical archive and ignores contemporary manifestations of apartheid’s racial logic. The Court can then avoid grappling with the full extent of injustice that arises from living in a racialised space in post-apartheid South Africa. Traditional legal methodology burdens the applicants with proving the pertinent aspects of their claim, in this case the adverse consequences of apartheid geography and its contemporary manifestations. Accordingly, the status quo is just unless shown otherwise. This creates a legal narrative in which race-based spatial disadvantage has ceased to exist through the elimination of discriminatory laws. This rewriting of the legal narrative erases intergenerational disadvantage faced by the residents of black ghettos.\textsuperscript{76} This narrative obscures the fact that the production of space is not solely driven by discriminatory law but is being reproduced despite race-neutral laws. By adopting this view, the Court casts Soweto’s space as neutral while effectively allowing spatial inequities to persist.

The above extract reveals another important aspect of the judgment: it illustrates the Court’s preference for a class-based frame.\textsuperscript{77} Curiously, in this case where racial discrimination was raised as a substantive ground for consideration, race or more precisely being a black inhabitant of a racialised space is not given much consideration by the Court. Instead, the Court prefers to frame the applicants as ‘poor’, ‘vulnerable’ or ‘desperate’.\textsuperscript{78} By avoiding the language of race, the Court obscures a key criterion not only in the creation of the township but a factor which continues to shape its occupants’ lives in contemporary South Africa. The fact of the matter is that the applicants were settled there because they were black.\textsuperscript{79} Given the country’s history of colonial and apartheid oppression, this subtle move towards a class frame diffuses one of the most powerful stratifying forces in the country’s history – race. To be sure, poverty

\textsuperscript{74} Ibid at para 150 (emphasis added).

\textsuperscript{75} Ibid.

\textsuperscript{76} Many of the residents of Soweto have developed as an urban ‘underclass’ of concentrated disadvantage. W Wilson \textit{The Truly Disadvantaged: The Inner City, The Underclass and Public Policy} (1987) 58; J Seekings & N Natrass (note 23 above) at 271.

\textsuperscript{77} S Sibanda ‘Now We’re Just “the Poor”: Race Consciousness and the Discourse of Socio-Economic Rights’ (unpublished work on file with author).

\textsuperscript{78} The judgment contains numerous instances in which the Phiri community is referred to as ‘poor’, while other references to the community in question speak of them being the ‘desperate’, ‘vulnerable’, or ‘those most in need’.

\textsuperscript{79} The shortcomings of this class-based narrative are exposed in the judgment of Froneman J in \textit{Daniels v Scribante} who illustrates that class was not the dominant determinant in the allocation of space. He points to state-led efforts to lift white people out of poverty which ignored black people. \textit{Daniels} (note 1 above) at para 132.
remains racialised, however, the manifold effects of racialisation cannot be reduced to purely economic considerations.80

As I have argued, apartheid geography extends beyond material deprivation to speak to the subordination of black people in all aspects of political and social life. This phenomenon is best captured by a structural understanding of racism according to which racism is instilled in systems and institutions which reproduce white supremacy despite the absence of overt acts of discrimination.81 The utility of this approach is that it moves the questions about racism away from overt acts or offensive laws but calls us to examine the substructure of society in which racism is embedded. In this view, it becomes impossible to characterise the residents of Phiri as targets of past discrimination. While this is true, it fails to capture the contours of their contemporary oppression. While I do not suggest that the Court avoids race completely (racial discrimination is a substantive ground behind the residents’ challenge), the Court’s approach to race often amounts to ‘noticing but not considering race’.82 Gotanda describes the technique of racial non-recognition in judicial analysis as follows:

> It addresses the question of race, not by examining the social realities or legal categories of race, but by setting forth an analytical methodology. This technical approach permits Courts to describe, to accommodate, and then to ignore issues of subordination. This deflection from the substantive to the methodological is significant. Because the technique appears purely procedural, its normative, substantive impact is hidden.83

This methodology despite its apparent neutrality effectively suppresses racial subordination – the socio-political cost of which is evident in limiting tools of resistance against racial discrimination. While the slippage between race and class may seem innocuous, it is anything but as Lopez argues:

> The language of race may well constitute the single most indispensable tool for combating and ameliorating the deleterious effects of racism. Because race is so deeply embedded in this society, its reach and effects must be addressed in terms of race itself—there is no better, indeed, no other language available to us.84

Despite the Court’s apparent reluctance to engage with race and space in this case, it was compelled to consider apartheid’s racist policy in order to address the issue of unfair discrimination on the grounds of race as this ground was explicitly raised by the applicants. Unsurprisingly, given the downplaying of the effects of race, the Court held that there was no such discrimination. Thus, in a few swift moves the Court dismisses the challenge to racialised spatial disadvantage by, first, emphasising the break from the past. Secondly, the Court locates the effects of apartheid geography in the historical archive and finally, the Court introduces

80 Pierre De Vos argues ‘escaping poverty and joining the middle class does not — as some have argued — free “black” South Africans from the effects of racial identity and race-based thinking’ in P De Vos ‘The Past is Unpredictable: Race, Redress and Remembrance in the South African Constitution’ (2012) 129 South African Law Journal 73, 78.


83 Ibid at 17.

an alternative narrative – the deracialised discourse of poverty. This approach leaves the racial configuration of space, power and privilege unquestioned.

Adopting a critical legal geographic approach and reading the space in this case reveals that racial oppression is written into the space’s very fabric. This partly explains the distinction made between Phiri and other areas which did not raise objections to the state’s policy. The discrimination in effect cannot be reduced to a single incident. While the Court frames the applicants as poor occupants of a historically disadvantaged space, weaving in insights from legal geography has the potential to enhance the accuracy of the applicants’ characterisation. Overlaying a socially contingent view of Phiri reveals that its composition is a manifestation of discrimination in and of itself. In this view, we can understand the applicants as subject not only to class-based disadvantage, but also to racial and spatial oppression. This composite view of discrimination makes the applicants appear infinitely more vulnerable as their marginalisation is shown to be imbedded in and reinforced by, not only their identity, but also their location in a racialised ghetto. This re-characterisation of the applicants could have influenced the Court’s ultimate reasoning in this case and, in the future, improve the outcomes for those similarly placed in legal proceedings.

B Daniels and the legal landscape of conquest

If Mazibuko shows the Court limiting its consideration of extraneous factors in spatial adjudication, the judgment of Madlanga J in Daniels marks a significant move towards recognising the impact of apartheid geography in legal adjudication. This case concerns the rights of occupiers to make improvements to their dwellings in terms of the Extension of Security of Tenure Act 62 of 1997 (ESTA). In my view, this case begins to explore the potential of a spatially contextualised approach to adjudication by interweaving the effects of racialised geography and a substantive reading of the Bill of Rights. In contrast to Mazibuko’s peri-urban setting, this case is drawn from the rural context which is indelibly marked by apartheid geography.

In this case, the Court was called on to consider another aspect of apartheid geography’s contemporary manifestation. The dispute arose over improvements to basic amenities Ms Daniels sought to make. These included ‘levelling the floors, paving part of the outside area and the installation of an indoor water supply, a wash basin, a second window and a ceiling’. The trouble began as Ms Daniels was an occupier of land owned by a property company and managed by Mr Scribante, who actively sought to remove Ms Daniels from the property through a range of measures calculated to make her continued occupation of the property untenable. These efforts included cutting the electricity supply to the property and failing to maintain the building to a habitable standard, such that Ms Daniels had to approach the Magistrates Court to direct Mr Scribante to restore the electricity supply and make repairs to the building.

The central legal question before the Court was whether an occupier in terms of ESTA was entitled to effect improvements to her dwelling without the owner’s consent. The case reached the Court on appeal after Ms Daniels saw her application dismissed with costs in the Stellenbosch Magistrates Court. Her appeal to the Land Claims Court was similarly unsuccessful: the court denied the existence of such a right, holding that it was too great an

---

85 Daniels (note 1 above) at para 7.
intrusion on the rights of owner to be granted save with the express inclusion in ESTA. Indeed, no such right is contained within the text of ESTA and the courts a quo were not prepared to recognise such a right. In adjudicating the dispute, the courts a quo considered the matter a legal question the answer to which lay purely in black-letter law. The absence of an express right to effect improvements meant the absence of remedy for Ms Daniels. Both the Land Claims Court and the Supreme Court of Appeal refused Ms Daniels leave to appeal.

The Constitutional Court reached a different conclusion. There the matter turned on the right to dignity enjoyed by all occupiers under ESTA in terms of s 5(a) which provides: ‘subject to limitations which are reasonable and justifiable in an open and democratic society based on human dignity, equality and freedom, an occupier, an owner and a person in charge shall have the right to human dignity’. Recognising the importance of dignity under ESTA’s legislative scheme, the Court premised its approach on determining what the right to dignity meant for Ms Daniels. The Court also relied on s 6(1) of the Act which provides that occupiers shall enjoy the right to ‘reside on and use the land on which he or she resides’ and ‘to have access to such services as had been agreed upon with the owner or person in charge, whether expressly or tacitly’.

The Court reasoned that the right to reside and have access to services cannot be reduced to mere residence in abhorrent conditions. The right to reside must be understood in light of the fundamental rights of occupiers of which dignity is central. According to the Court ‘occupation is not simply about a roof over the occupier’s head… it is about more than just that. It is about occupation that conduces to human dignity and the other fundamental rights’. Accordingly, Ms Daniels was entitled to make basic improvements to her home in order to bring it to a standard befitting of a dignified life. Mr Scribante raised the point that nowhere in the Act is provision made for a right to make improvements. The Court dismissed this argument holding that if the rights of occupiers were to be so narrowly interpreted such that they were limited solely to those itemised in the Act, this would render the other rights, the context and the purpose of the Act hollow. The Court cautioned against adopting a narrow technical interpretation which would restrict the occupiers’ rights. The Court found that a holistic interpretation was therefore necessary:

A denial of the existence of the right asserted by Ms Daniels might inadvertently result in what would in effect be evictions. This would be a direct result of the intolerability of conditions on the dwelling. And these “evictions” might happen beneath the radar of the carefully crafted eviction process. That would make nonsense of the very idea of security of tenure. After all, like the notion of “reside”, security of tenure must mean that the dwelling has to be habitable. That in turn connotes making whatever improvements that are reasonably necessary to achieve this. Of what use is a dwelling if it is uninhabitable? None.

The Court thus overturned the decision of the courts a quo holding that Ms Daniels did in fact have the right to make improvements to her dwelling. In arriving at its conclusion, the Court looked beyond the relevant provisions, the black letter law and the submissions made by the parties. In a bold step, the Court, of its own volition, adopted a frame grounded in

---

86 ESTA, s 5(a).
87 Ibid, s 6(1).
88 Daniels (note 1 above) at para 31.
89 Ibid at para 26.
90 Ibid at para 29.
91 Ibid at para 32.
the narrative of black conquest and dispossession to locate this case in its spatial context. In so doing, the Court read the context temporally as well as spatially. This reshapes the nature of the dispute from the rights of ESTA occupiers to improve their dwellings, to a case about the restoration of dignity for those holding precarious land rights due to apartheid geography. This subtle re-characterisation is powerful as it gives voice to a discourse often marginalised in the resolution of spatial disputes. It allows the Court to delve into the production of space with striking clarity on the racialised patterns of spatial development showing that the rural context in which this case is located is an embodiment of apartheid geography. The Court achieves this by unpacking the human geography which underpins the dispute finding that:

Dispossession of land was central to colonialism and apartheid... The purpose of it all was, first, the obvious one of making more land available to white farmers. The second “was to impoverish black people through dispossession and prohibition of forms of farming arrangements that permitted some self-sufficiency. This meant they depended on employment for survival, thus creating a pool of cheap labour for the white farms and the mines. White farmers had repeatedly complained that African people refused to work for them as servants and labourers”. The third was the enforcement of the policy of racial segregation, which assumed heightened proportions during the apartheid era.

By revealing the full historical context of the space in question, the Court departed from its narrower approach to space evident in *Mazibuko*. This suggests that the Court was taking note of the manifestations of apartheid geography and reading the law in a space-conscious manner. Traditional legal methods rarely call for such nuanced contextualisation. Instead the focus tends to be on positive legal texts considered in the abstract.

Race was a key element of the Court’s chosen frame. The Court dedicated much of the judgment to tracing Ms Daniels predicament to its historical roots in the dispossession of black people of their land under colonialism and apartheid. This tracing reveals the racial production of space and the inequitable landscapes it has created. By highlighting the salience of race, the Court avoided casting the problem as being one of generalised poverty in order to make a sharper, more accurate point about the racial and spatial inequality at the root of the South African economic order. The Court then incorporated this narrative into the interpretation of the legal framework. In elaborating the milieu in which ESTA must be understood, the Court found that:

Apartheid sought to divest all African people of their South African citizenship. According to the grand scheme of apartheid, Africans were to be citizens of so-called homelands. The consequence was a variety of tenuous forms of land tenure for victims within what – to apartheid – was “South Africa proper”. This meant throughout the length and breadth of our country victims were made strangers in their own country. On farmland – which this case is about – their residence was particularly precarious. They could be, and were often, subjected to arbitrary evictions. Needless to say, they could not have much say on the conditions under which they lived on the farms, however deplorable. This was a life bereft of human dignity.

This illustrates the Court’s openness to expanding the range of relevant considerations. The Court made it clear that while the ‘legal’ question may have been presented by the parties narrowly, the Court was prepared to go beyond a purely legal view towards a framing which

92 Ibid at paras 13–58.
93 Ibid at para 14.
94 Ibid at para 16.
95 Ibid at para 21.
reflected the reality as experienced by Ms Daniels and countless others. This included the country’s history of racialised forced removals and spatial inequality. In this approach, the manifestations of apartheid geography are not only explicitly named but sought to be addressed through reliance on an expanded and contextualised reading of the Constitution’s human rights framework. This approach alters dominant representations of space and builds the resources necessary to remediate some of the violence of apartheid geography. The Court was deliberate in outlining the racial and spatial policies which produced the applicant’s predicament. Importantly, this allowed the Court to read the entire socio-spatial context and not merely the legal text as the courts a quo had done. Reading the social and spatial context moved the analysis away from the letter of the law towards considering the racial and spatial power dynamics which underpin the case’s spatial context. By interrogating the production of space, the Court helped to disrupt the power relations which reinforce spatial inequality. Thus the background of spatial racism was brought into the foreground of the case. This enriched view of law, space and society (including race and gender-based discrimination) allowed the Court to reach a conclusion which would have been unsupported by a mere technical reading of law. This, along with the basic nature of the upgrades Ms Daniels sought to make, tipped the balance in her favour but also expanded protections for similarly positioned occupiers. Beyond Ms Daniels’ circumstances, this approach holds immense potential to recalibrate the dominant understanding of space and to reshape laws in favour of substantive remedies for those faced with insecure access to land and housing rights in post-apartheid South Africa.

While the narrative in Mazibuko suggests a break with the past, the Court in Daniels is clear that the historical context is not always confined to the past. Drawing clear connections between the past and the present, the Court observed:

Painfully, in some instances this [legacy of land dispossession] is not just history. To this day, some of the poorest in our society continue to keep homes under the protection of ESTA. Needless to say, occupiers under ESTA are a vulnerable group susceptible to untold mistreatment. This is especially so in the case of women.96

In this extract we see the Court drawing connections between time and space not only as an introduction to an atemporal or decontextualised legal question, but also as a link between history and contemporary spatio-legal problems. The benefit of this approach is that it does not misrepresent the conditions in the country’s spaces as being instantly transformed by constitutional change. Instead, it creates a link with the past which highlights the necessity for effective remedies. The majority’s approach thus moves towards recognising the current effects of apartheid geography and not merely its historical existence. The Court achieves this by first outlining the racial and spatial context which gives rise to the dispute. Secondly, the Court interprets the legal framework bearing in mind the context and what the law seeks to achieve. Finally, the Court arrives at a determination that best responds to the injustice caused by the context of apartheid geography.

In a separate concurring judgment, Froneman J explored the continuity with the past to disrupt the dominant narrative about the spatial, social and economic status quo in South Africa. He argued that: ‘historical injustice is nowadays not easily denied, but rather avoided’.97

This call to revisit even norms that appear neutral comes after he had argued that even these conditions had been socially engineered in a historical narrative which saw the state drive

---

96 Ibid at para 22.
97 Ibid at para 117.
inequality in favour of white people. He observed that history is often suppressed by discourses of racial difference, inferiority and misfortune which seek to provide an explanation for the lowly position of many black people in society. He argues:

[the] Carnegie Commission’s report on poor whites showed that poverty, also on farms, had nothing to do with inherent inferiority, but everything to do with social and economic processes outside individual control. The problem was addressed so that white people could maintain dignified living standards. The burning injustice, namely that this corrective action was not extended to black and “coloured” people, must and can be rectified.98

This remedial imperative is of course far broader than can be addressed by a single judgment. However, Froneman J expanded on the contribution of the majority judgment. He argued that it had to be understood in the context of the changing nature of property relations that the Court has sought to remedy in a series of property and housing-related judgments.99 The approach adopted by the majority, and expanded on by Froneman J, certainly disrupted the judicial orthodoxy. The Court’s bold retelling of a suppressed narrative and astute reframing of a legal dispute contributes to the resources on which judges and parties may draw. This potential is, however, tempered by Cameron J’s separate concurring judgment which, while approving of the majority’s approach, alluded to the dangers of judicial pronouncements on history. This judgment set out some of the limits in the majority’s approach. Cameron J cautioned that:

It is not within the primary competence of judges to write history. The histories in my colleagues’ judgments were not expressly in issue during argument before us. Neither side referred to them. We did not have the benefit of the parties’ contesting approaches to or submissions on them. And the parties placed before us none of the historical sources my colleagues refer to and quote from. This means we are on spongy ground. And we could lose our step. Especially where accounts are incomplete and where they are not directly functional to the determination of the dispute.100

This cautionary dictum partly explains the Court’s difficulties in addressing apartheid geography. Given that the challenges presented by apartheid geography are rarely purely legal, can the constitutional matrix, premised as it is on judicial review, respond and remediate the ills of racialised space? Although Ms Daniels’ case can be viewed in isolation, she is not alone. Ms Daniels represents countless occupiers who pursue a dignified existence in a milieu characterised by violent dispossession and landlessness. In this environment, private and state-sanctioned violence are normalised as rights and the lack thereof reinforce a range of responses. By reading the narrative of racial and class-based inequality into inquiries about what rights are enjoyed by occupiers and owners, we can give meaning to the promise of dignity and move the law towards achieving equality. This can be achieved by expanding the reading of legal facts and concepts, such as the ‘(unlawful) occupier’, in favour of interpretations which appreciate the interplay of race, space and other stratifying categories. By overlaying social and spatial facts in the dispute, the Court’s approach in Daniels resonates with a critical legal geographic reading of law. This methodology has the potential to increase positive outcomes for those adversely affected by apartheid geography.

98 Ibid at para 132.
99 S Wilson (note 27 above) at 280.
100 Daniels (note 1 above) at para 149.
IV CONCLUSION

Addressing the pernicious effects of colonialism and apartheid lies at the heart of South Africa’s constitutional project. The success or failure of this project will hinge on its ability to heal the wounds of racial oppression. These wounds are painfully evident in the country’s divided spaces. This article has sought to analyse the operation of law in the context of a central tenet of apartheid geography. In so doing, I explored the landscape of apartheid and some of its oppressive manifestations. This apartheid geography, I argue, continues to have profound effects on social and political life across rural and urban settings. This article then introduced critical spatial theory into South Africa’s legal discourse as a way of understanding some of the manifestations of apartheid geography. The overall aim is to move legal discourse towards an understanding of space which is grounded in subjects’ actual socio-legal experiences. This moves the discourse towards a contextualised mode of legal analysis, which comes closer to reflecting everyday life. Through the lens of critical legal geography, I juxtaposed Mazibuko and Daniels – important cases in the canon of post-apartheid spatial jurisprudence. This revealed that while mentioning apartheid geography is almost unavoidable, legal constructions of space in this regard are contested and varied.

Evident in the Mazibuko approach to race is the tendency to notice the claimants’ race and space in deciding the case. However, this amounts to ‘noticing but not considering’ the full extent of the subjects’ subordination. This approach subtly downplays part of the subjects’ plight by not examining the social realities in which the claimant is located. Thus, the treatment of racialised space in Mazibuko falls short of an engagement designed to remediate the effects of racial classification, segregation and domination. This non-recognition of the historical and contemporary manifestations of spatial inequality produces harm while courts are seemingly trying to right wrongs. For example, by casting legal subjects merely as ‘unlawful occupiers’, legal analysis fails to recognise that the spatial is a key determinant of the subjects’ history, social origin and identity. Failing to capture the effects of geography in legal terms, limits the law’s ability to provide meaningful redress. A more context-driven understanding of law and space is required to adequately capture the harm faced by the subjects of racialised space. Ultimately, ‘this is a critique of a theoretical tradition, long dominant in the social sciences, of treating context as a container in which – but not because of which – important things happen’. I argue that, the Court in Daniels moves to address many of these shortcomings in what is a pioneering judgment. The Court draws on history to recognise that Ms Daniels’ predicament arose because of her location within a space marred by apartheid geography. This frame introduces into legal normality an idea which users of space appreciate intuitively – space matters. The potential of this approach lies in its ability to highlight socially-embedded power imbalances which are mirrored in spatial relations. This, I argue, will allow for better remedies and produce beneficial results for law and policy on space. In this context, the majority judgment in Daniels marks a welcome move towards a substantive understanding of apartheid geography.

A Service Conception of the Constitution: Authority, Justification and the Rule of Law in Proportionality Jurisprudence

RICHARD STACEY

ABSTRACT: Constitutional Court judgments upholding or striking down statutory provisions that limit constitutional rights in South Africa do so on the back of an inquiry into the proportionality of the rights limitation. The Court’s record suggests that proportionality analysis usually proceeds along one of two paths: an analysis that focuses on the availability of means less restrictive of rights than the impugned limitation, or an analysis that considers the strict balance between the value of the purposes sought by the limitation and the seriousness of the rights limitation. This article examines how these two paths of proportionality analysis appear at first to align with two different conceptions of the rule of law. Strict-sense proportionality analysis involves reasoning from moral first principles – or primary moral reasoning – that draws on Lon Fuller’s inner morality of law, and which seeks to provide morally grounded justifications for a decision upholding or striking down a limitation. Less restrictive means analysis depends on a ‘service conception’ of the constitution to justify its conclusions, drawing on Raz’s views about how law and legal decisions claim authority. Both Raz’s service conception of authority and the service conception of the constitution that runs through less restrictive means analysis, however, depend on a view of what it means to be a legal subject that is ultimately indistinguishable from Fuller’s: on either account, the rule of law demands that a legal system recognise its subjects as morally autonomous agents deserving of morally intelligible explanations for the law’s operation. If less restrictive means analysis is to make a meaningful claim both to authority and to the rule of law, then it must – and indeed does – involve at least some primary moral reasoning. In turn, South Africa’s culture of justification demands that courts embrace primary moral reasoning in proportionality jurisprudence.

KEYWORDS: moral reasoning, limitations, restrictive means, balancing, Fuller, Raz

AUTHOR: BA LLB (Wits) PhD (NYU). Associate Professor, University of Toronto Faculty of Law. Email: richard.stacey@utoronto.ca

ACKNOWLEDGEMENTS: An early draft of this article was presented at the Constitutional Court Review IX conference, 2–3 August 2018, Constitution Hill, Johannesburg. I am grateful to my fellow participants at the conference for discussion of the ideas presented here, and to the editors of this journal and its anonymous reviewers for comments on early drafts of this article.
I INTRODUCTION

South Africa’s 1996 Constitution, like others in the family of post-war, dignity-protecting constitutions, entrenches a handful of liberal democratic rights that can only be limited by legislation if it is reasonable and justifiable to do so.¹ In a seminal decision under South Africa’s Interim Constitution (1994–1996),² the Constitutional Court understood this inquiry into reasonableness and justifiability to require ‘the weighing up of competing values, and ultimately an assessment based on proportionality.’³ A few years later, the Court described the analysis as ‘a balancing exercise’ that must ‘arrive at a global judgment on proportionality,’ and which demands that ‘the more serious the impact of the measure on the right, the more persuasive or compelling the justification must be’.⁴

The principle of proportionality has become the analytical fulcrum of limitations jurisprudence in South Africa, as it has in other jurisdictions with similar constitutional limitations clauses.⁵ In this article I explore two paths along which the pursuit of proportionality has proceeded in South Africa’s constitutional jurisprudence since the 1990s, two paths which track a broader debate in the literature about how best to engage in proportionality analysis. Much more than this, though, I argue here that the points of intersection between these paths of proportionality allow us to make important observations about the nature of law, authority and justification in a contemporary constitutional democracy.

One of these paths involves the balancing exercise itself, or proportionality ‘in the strict sense’. The inquiry on this approach goes to whether the objective a rights-limiting measure aims to achieve is important enough to justify the rights limitations that the measure imposes. On balance, limiting a right is justifiable if it can be demonstrated that the value of what is achieved by doing so outweighs the harm it causes. As I have argued elsewhere, however, the balancing metaphor is misleading: an analysis of proportionality in its strict sense should not be conceived of as the direct weighing up of the inherent value of rights and competing statutory objectives, but rather as asking whether upholding a right or allowing its limitation in

---

¹ Section 36(1) of the Constitution of the Republic of South Africa, 1996 (Constitution), reads: 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.
² Constitution of the Republic of South Africa, Act 200 of 1993 (Interim Constitution). The Interim Constitution has been repealed.
⁴ S v Matamela & Another [200] ZACC 5, 2000 (3) SA 1 (CC) at para 32. See also S v Bhulwana; S v Gwadiso [1995] ZACC 11, 1996 (1) SA 388 (CC) at para 18 (‘The more substantial the inroad into fundamental rights, the more persuasive the grounds of justification for the infringing legislation must be.’)
pursuit of competing objectives better advances a society’s commitment to a set of normative principles or values.\(^6\)

The second path of proportionality analysis asks the more fact-centred question of whether the legislature has selected means to achieve its objectives that are the least restrictive or are minimally invasive of constitutional rights. The ‘least (or less) restrictive means’ test understands the concept of proportionality to require statutory provisions to limit constitutional rights no more than is necessary to achieve important objectives. If there are less restrictive means available to achieve a given objective, the conclusion must be that the statutory provision is more restrictive than it needs to be, and hence is disproportionate.

The South African Constitutional Court – and many other high courts around the world – considers both the availability of less restrictive means and proportionality in the strict sense as elements of a global inquiry into proportionality. But the Court often relies more heavily on one inquiry than the other, and there are cases in which the Court relies exclusively on one or other inquiry in coming to a conclusion on the proportionality (and thus the reasonableness and justifiability) of a rights limitation.\(^7\)

In parts II and III of this article I look at the South African Constitutional Court’s decisions in *Lawyers for Human Rights v Minister of Home Affairs*\(^8\) and *Chevron v Wilson*,\(^9\) two recent and brief judgments emblematic of reasoning along these two paths of proportionality. The brevity of the proportionality analysis in each judgment indicates just how established each mode of reasoning has become, with the Court content to do little more than gesture at either mode of reasoning in order to bring its persuasive force to bear on the resolution of a particular dispute. However, a court’s role in rights limitations cases is not only to determine whether a particular rights limitation is justifiable or not, but also to explain why that limitation is justifiable or not. A commitment to the rule of law requires that the reasons for law’s application, for court judgments just as for any other legal rule, be intelligible to the people bound to obey it. And in turn, a judgment’s claim to authority – that is, its claim to stand as an authoritative statement of law that makes a demand on legal subjects’ obedience – depends on whether its reasoning is intelligible to legal subjects.

I argue in part IV, then, that these two modes of reasoning about proportionality correspond to two apparently quite different conceptions of law in general, and of the rule of law in particular. An approach that foregrounds proportionality in the strict sense aligns closely with Lon Fuller’s account of the rule of law and the idea that law as such has an ‘inner morality’.\(^10\) Law’s authority on this Fullerian approach depends on its justifiability against the normative principles or values to which a political community happens to commit. Demonstrating the justifiability of a rights-limiting statutory provision – or its absence – is the job of courts of review, and a court’s conclusion on this issue will itself be authoritative to the extent that it

---


\(^9\) *Chevron SA (Pty) Ltd v Wilson t/a Wilson’s Transport [2015] ZACC 15, 2015 (10) BCLR 1158 (CC)(*Chevron v Wilson*).

intelligibly demonstrates that either upholding or striking down the limitation is consistent
with the legal system’s normative foundations.

On the other hand, judicial reliance on less restrictive means analysis presents what I call a
‘service conception of the constitution’, adapting Joseph Raz’s service conception of authority.11
Judgments that rely on the ostensibly objective, fact-based heuristic of the less restrictive
means analysis make a claim to authority, on this Razian view, because they produce decisions
about the validity or invalidity or rights limitations that are more intelligibly related to our
most basic normative commitments than when judges engage in unavoidably subjective and
impressionistic moral reasoning about what those commitments require. On this view, the
moral reasoning that strict-sense proportionality analysis is abhorrent to the rule of law, and
should be avoided, because it is unintelligible to legal subjects (or at least, less intelligible than
less restrictive means analysis).

It is interesting on its own to identify the different conceptions of law and of the rule of
law at work in each of these modes of proportionality analysis. But the more interesting claim
I make is that Raz’s service conception of authority, and with it the service conception of the
constitution that runs through less restrictive means analysis, depends on a view of what it
means to be a legal subject that is ultimately not all that different from Fuller’s.12 Legal subjects,
I argue along with Fuller, demand justification for the operation of the legal system that goes
beyond purely factual reasoning of the kind that less restrictive means analysis purports to
deliver, and this is no less the case with respect to rights limitations. This follows from the
law as such, for Fuller, while for Raz it follows from a merely contingent commitment to the
rule of law.

The implications for these two paths of proportionality are significant. Critics of strict-sense
proportionality analysis prefer less restrictive means analysis for the reason that it provides
more intelligible reasons for law’s operation, and thus better upholds the rule of law. But if a
commitment to the rule of law demands that the legal system recognise its subjects as morally
autonomous agents deserving of morally intelligible reasons explaining why the law operates
as it does – as I argue it does for Raz as much as for Fuller – then less restrictive means analysis
cannot both uphold the rule of law and avoid moral reasoning. It follows, as I argue in part V,
that if less restrictive means analysis is to make a meaningful claim both to authority (ie, to
producing legal judgments that we should respect and obey) and to the rule of law, then it
must – and indeed does – involve at least some reasoning from moral first principles (what I
refer to here as ‘primary moral reasoning’).13 For South Africa’s legal system to continue to be
a ‘culture of justification’,14 we must acknowledge that both paths of proportionality analysis
depend on and demand the kind of justification that only moral reasoning can deliver, and
which proportionality in the strict sense unashamedly involves.

12 K Rundle ‘Form and Agency in Raz’s Legal Positivism’ (2013) 32 Law and Philosophy 767 and K Rundle Forms
Liberate: Reclaiming the Jurisprudence of Lon L Fuller (2012).
13 R Stacey ‘The Magnetism of Moral Reasoning’ (note 6 above). See also T Hickman ‘Proportionality:
Comparative Law Lessons’ (2007) 12 Judicial Review 31, 47; M Zion ‘Effecting Balance: Oakes Analysis
Human Rights 31, 32. See also M Cohen-Eliya & Iddo Porat ‘Proportionality and the Culture of Justification’
II LAWYERS FOR HUMAN RIGHTS V MINISTER OF HOME AFFAIRS

A The normativity of ‘global’ proportionality analysis

Section 36’s proportionality inquiry instructs courts to consider not only whether the limitation of a right is ‘reasonable and justifiable in an open and democratic society based on dignity, equality and freedom’, but also whether it satisfies five other distinct desiderata. These narrower and more formal criteria embrace the nature of the right (section 36(1)(a)), the importance of the purpose of the limitation (section 36(1)(b)), the nature and extent of the limitation (section 36(1)(c)), the relation between the limitation and its purpose (section 36(1)(d)), and less restrictive means to achieve the purpose (section 36(1)(e)).

The inquiries in this analysis largely correspond to what has become a standard judicial approach to limitations analysis in jurisdictions with similar constitutional limitations clauses. In Canada and Germany, for example, the high courts have set out a rubric for proportionality analysis that includes an initial inquiry into whether a rights-limiting measure pursues an important objective (respectively described in those countries as a ‘pressing and substantial’ objective and a ‘legitimate purpose’), two more factual inquiries into whether the limitation is rationally connected to that objective and whether there are less restrictive means to achieve that purpose, and finally the more evaluative question of whether the extent of the rights limitation is proportional to the benefits it produces.15

The text of s 36 of the South African Constitution reflects all four elements of this model of proportionality analysis, but the Court frequently recites the incantation that s 36 requires a global or holistic assessment of proportionality driven by the values at the heart of the Constitution. Consider, for example, Sachs J, writing in a separate concurring judgment:

The notion of an open and democratic society is thus not merely aspirational or decorative, it is normative, furnishing the matrix of ideals within which we work, the source from which we derive the principles and rules we apply, and the final measure we use for testing the legitimacy of impugned norms and conduct. If I may be forgiven the excursion, it seems to me that it also follows from the principles laid down in *Makwanyane* that we should not engage in purely formal or academic analyses, nor simply restrict ourselves to ad hoc technicism, but rather focus on what has been called the synergetic relation between the values underlying the guarantees of fundamental rights and the circumstances of the particular case.16

Perhaps more forcefully, a unanimous Court had this to say in *Manamela* in 2000:

It should be noted that the five factors expressly itemised in s 36 are not presented as an exhaustive list. They are included in the section as key factors that have to be considered in an overall assessment as to whether or not the limitation is reasonable and justifiable in an open and democratic society. In essence, the Court must engage in a balancing exercise and arrive at a global judgment on proportionality and not adhere mechanically to a sequential check-list. As a general rule, the more serious the impact of the measure on the right, the more persuasive or compelling the justification must be. Ultimately, the question is one of degree to be assessed in the

---

15 In Canada, the leading cases that emphasise this approach are: *R v Oakes* [1986] 1 SCR 103 at paras 69–70; *R v Big M Drug Mart Ltd* [1985] 1 SCR 295 at para 139; and *Canada (Attorney General) v JTI-Macdonald Corp* [2007] 2 SCR 610 at para 36. In Germany, the proportionality test was developed in *Nudism Education*, Bundesverfassungsgericht [BVerfG] [Federal Constitutional Court] 1958 7 Entscheidungen Des Bundesverfassungsgerichts [BverfGE] 320; *Pharmacy* BVerfG 1958 7 BVerfGE 377; *Own Account Transport Tax* BVerfG 1963 16 BVerfGE 147; and *Shop Closing Act II* BVerfG 1961 13 BVerfGE 237.

concrete legislative and social setting of the measure, paying due regard to the means which are realistically available in our country at this stage, but without losing sight of the ultimate values to be protected.  

These passages suggest that the Court is committed to a mode of reasoning about proportionality that draws connections between right and limitations on one hand and the fundamental normative commitments the Constitution makes on the other hand. ‘Purely formal’ analysis, ‘technicism’, or adhering ‘mechanically to a sequential check-list’ should support but not replace this primary inquiry into whether rights limitations can be persuasively justified against the Constitution’s most basic values. In practical terms, the Court’s decision-making proceeds on the principle that more serious infringements of rights, where ‘seriousness’ is understood in light of constitutional values, must be justified by more compelling reasons – where the force of those reasons is again assessed in light of basic values. On the face of these statements and the Court’s frequent reference to or recitation of them, it would seem that proportionality in the strict sense is the touchstone of South Africa’s limitations jurisprudence.  

At most, though, proportionality in the strict sense is solely determinative in only some of the Court’s decisions. Niels Petersen takes the view that proportionality in the strict sense plays almost no role in the Constitutional Court’s record of limitations decisions, finding only four cases in a sample of 44 to rely on a finding that the limitation was strictly disproportional to strike it down.  

My view is that proportionality in the strict sense plays a role in somewhat more than just four of the Constitutional Court’s rights cases, but the trifling difference between my reading of the cases and Petersen’s is of no concern here.  

Rather, the point is that there is a handful of cases, however large that handful actually is, in which the Court’s determination of a rights limitation’s justifiability depends on proportionality in the strict sense rather than on the formal and factual inquiries into rational connection and less restrictive means.

B Strict proportionality in Lawyers for Human Rights

Lawyers for Human Rights v Minister of Home Affairs did not involve a particularly challenging set of facts, and the judgment does not abound with ground-breaking judicial reasoning. It was on the contrary a quite straightforward case, involving a challenge to the constitutionality of sections of the Immigration Act 13 of 2002 on the grounds that they unjustifiably infringed
constitutional protections against detention without trial and arbitrary deprivation of freedom.\footnote{Constitution, s 12(1)(a) and (b) reads, in relevant part:
Everyone has the right to freedom and security of the person, which includes the right–
(a) not to be deprived of freedom arbitrarily or without just cause;
(b) not to be detained without trial …}
The Constitutional Court agreed with the Pretoria High Court that the impugned sections did indeed limit these rights and that these limitations were not reasonable and justifiable, grounding its decision in a broad application of strict-sense proportionality reasoning.

The Court began by noting that the objective of the Immigration Act is to regulate the admission to and departure of foreigners from South Africa. Its preamble declares that the detection, reduction and deterrence of illegal immigration are objectives of the Act, while also acknowledging that the contributions of migrants to South Africa are valuable and that it is important to promote human rights and minimise xenophobia. Much of the scheme of the Act, however, is geared towards finding and deporting illegal migrants. Section 2(1)(c), for example, provides that one of the primary objectives of the Department of Home Affairs is ‘detecting and deporting illegal foreigners’. Section 1 then defines an ‘illegal foreigner’ as a person who is neither a citizen nor a permanent resident and is in contravention of the Act or is a ‘prohibited person’, and s 29 in turn defines prohibited persons broadly, to include people sick with infectious diseases, people with criminal convictions, and people who have previously been deported.\footnote{Section 29(1)(Prohibited Persons)(The following foreigners do not qualify for a temporary or a permanent residence permit:
(a) those infected with infectious diseases as prescribed from time to time;
(b) anyone against whom a warrant is outstanding or a conviction has been secured in the Republic or a foreign country with which the Republic has regular diplomatic relations in respect of genocide, terrorism, murder, torture, drug trafficking, money laundering or kidnapping;
(c) anyone previously deported and not rehabilitated by the Department in the prescribed manner;
(d) a member of or adherent to an association or organisation advocating the practice of racial hatred or social violence; and
(e) anyone who is or has been a member of or adherent to an organisation or association utilising crime or terrorism to pursue its ends.)}

To reduce and deter illegal migration, the Act includes provisions that grant wide-ranging powers to authorities to detain migrants for the purposes of deporting them. Section 34 of the Act, the provision at the heart of the case, empowers immigration officers to arrest ‘illegal foreigners’ without a warrant and detain them ‘at a place determined by the Director-General’ of the Department of Home Affairs until they are deported. Section 34(1)(d) provides that such persons may be detained for up to 30 days without a warrant of court, and for up to 90 days thereafter with a warrant of court. As the Constitutional Court put it, the section ‘grants drastic power to an administrative official’.

The applicants’ primary complaint against the provision was that it imposed neither constraints nor guidelines on officials’ exercise of this drastic administrative power, nor provided an opportunity for judicial scrutiny until after 30 days of detention. And even for detention beyond 30 days, section 34(1)(d) did not require the court considering the Department’s application for a warrant to hear the detainee in person. Quoting from its 1998 discussion of the prohibition on detention without trial in \textit{De Lange v Smuts}, the Court noted:
History nevertheless emphasises how important the right not to be detained without trial is and how important proper judicial control is in order to prevent the abuses which must almost inevitably flow from such judicially uncontrolled detention.

Having decided that the impugned provisions limited constitutional rights, the Court spent a few short paragraphs determining that the limitations were not reasonable or justifiable and could not therefore be saved by s 36 of the Constitution. The state sought to justify the right’s limitation on the basis that it would prove costly and judicially onerous to bring every detained migrant before a court. The objective served by limiting illegal migrants’ fair trial and freedom rights, the state argued, was avoiding the costs that would be incurred by respecting those rights.

In responding to this argument, the Court did not bother with a sequential procession through the inquiries into whether saving costs is an important or legitimate objective by itself, if the rights-limiting measure was rationally connected to that objective or whether there were less restrictive means available to achieve it. Rather, the Court satisfied itself with what seems to be an uncomplicated conclusion that the cost-saving benefits of limited judicial oversight of immigration detention do not justify the harms that flow from limiting rights to fair trial and personal freedom. The Court found the state’s arguments in this respect ‘woefully short of justifying the limitations’.

This is an example of proportionality analysis in its strict sense, with no reliance at all on whether less restrictive means to deter illegal migration were available. Even though the Court could have considered whether it would be less invasive of rights to fair trial to require a warrant for any detention beyond, say, 48 hours or 72 hours rather than 30 days, it did not do so. An inquiry into less restrictive alternatives was not necessary to the Court’s conclusion that the measure was disproportionate. Rather, the Court’s decision relies on the assessment that it is more important, as a matter of constitutional principle, for the law to prohibit the limitation of rights to fair trial and personal freedom than it is for the legal system to save the costs of additional court hearings.

Lawyers for Human Rights lies along a path of jurisprudence in which strict-sense proportionality analysis figures large in the resolution of limitations inquiries. This jurisprudence has established strict-sense proportionality analysis as a recognised form of reasoning in constitutional litigation, and perhaps as a result, the reasoning in Lawyers for Human Rights is

23 Ibid at paras 52–58.
24 Ibid at para 59.
25 Minister of Home Affairs v National Institute for Crime Prevention and the Reintegration of Offenders (NICRO) & Others [2004] ZACC 10, 2005 (3) SA 280 (CC) at paras 47–50 (Court accepted that avoiding administrative burdens could be a legitimate objective for the state to pursue.) See also Prince v President of the Law Society of the Cape of Good Hope [2002] ZACC 1, 2002 (2) SA 794 (CC)(Court held that a religious exemption to the criminal prohibition on cannabis use, while a less restrictive alternative than a blanket prohibition, would impose additional costs and administrative burdens on the police that would undermine their ability to achieve the main objectives of reducing the harmful effects of narcotic drug use.) But see Minister of Justice and Constitutional Development & Others v Prince [2018] ZACC 30; 2018 (6) SA 393 (CC); 2019 (1) SACR 14 (CC)(Court holds that the blanket prohibition on dagga use violates the right to privacy in a manner that could not be justified by the law’s objective of reducing its allegedly harmful effects or complying with international treaties regarding drug trafficking.)
26 Lawyers for Human Rights (note 8 above) at paras 61 and 63.
not as full or as explicit as in some of those earlier cases. I refer to Lawyers for Human Rights here, not because it is an exemplar of clear and thorough strict-sense proportionality reasoning but precisely because it is not. The Court’s argument is merely emblematic of proportionality in the strict sense as a form of argument on which the Court can rely when appropriate. And because that form of argument is established and recognisable, the Court need do little more to justify its conclusion than to say that the reason for the limitation – saving costs – is not compelling or persuasive enough to warrant a serious limitation of rights to fair trial and personal freedom.

This path of reasoning has become so embedded in South African constitutional jurisprudence, such that the Court could follow it in somewhat pedestrian fashion in Lawyers for Human Rights, only because several judgments before now have already laid out in a deeper way how to think through questions justifiability in light of constitutional values. In the reference regarding S v Walters, for example, the Court concluded that shooting dead a person suspected of having committed a crime as petty as stealing an apple from a fruit-stand is a manifestly disproportionate way to pursue the objective of apprehending people suspected of having committed crimes.27 The Court’s analysis begins by noting that s 36 ‘requires a weighing-up of the nature and importance of the right(s) that are limited together with the extent of the limitation as against the importance and purpose of the limiting enactment’.28 Framing the inquiry explicitly in light of constitutional values, the Court went on:

What looms large … in the present case, is that the right to life, to human dignity and to bodily integrity are individually essential and collectively foundational to the value system prescribed by the Constitution. Compromise them and the society to which we aspire becomes illusory. It therefore follows that any significant limitation of any of these rights, would for its justification demand a very compelling countervailing public interest.29

The flaw in the statutory scheme that authorised the use of deadly force, the Court said, was that it drew no distinction between the use of such force in different situations. If a fugitive poses no threat to other people, it went on, there is no justification for the use of deadly force.30 Only the pursuit of objectives as compelling as the need to protect rights to life, dignity and bodily integrity more broadly – that is, objectives closely tied to the Constitution’s value system – could justify the use of deadly force.

Following similar logic, in S v Niemand the Court struck down statutory provisions requiring seven-year minimum prison sentences for ‘habitual criminals’, even for the commission of crimes ‘which do not constitute violence or a danger to society’. The Court held that the provisions limited the right not to be arbitrarily deprived of freedom, and further that the limitation was ‘grossly disproportionate’ to the value of separating non-violent criminals from society.31

More recently, the Court was asked to decide whether a common-law action in tort should continue to lie to a non-adulterous spouse against a third person with whom an adulterous spouse had engaged in an affair.32 The common-law wrong of contumelia, the Court noted, exists because it is a ‘blatant intrusion on the dignity of one partner to introduce a new member to that union without obtaining that partner’s consent’.33 The common-law action serves a

---

27 Walters (note 18 above) at paras 41–46.
28 Ibid at para 27.
29 Ibid at para 28.
30 Ibid at para 46.
31 Niemand (note 18 above) at paras 19 and 25.
cuckolded partner’s dignity by allowing them to recover damages from third persons, and protects the dignity of spouses more generally to the extent that it may deter adultery and preserve the sanctity of marriage. The crisp question before the Court was whether these objectives justified intrusions into both the adulterous spouse’s and the third person’s rights to privacy, ‘to have a sexual relationship with whomever he or she chooses’, and to personal freedom and security.

The Court decided that the action for *contumelia* will no longer be part of South African common law because it cannot be considered wrongful, in light of the ‘community’s general sense of justice’, the ‘legal convictions of the community’ or ‘public policy’, for one person to engage in a sexual relationship with another person who happens to be married. The Court made it clear that the determination of wrongfulness depended on the ‘balance’ to be struck between the potential infringement of non-adulterous spouses’ dignity on one hand, and the actual infringement of adulterous spouses’ and third parties’ rights on the other hand. Moreover that balance – and thus wrongfulness in this case – is ‘informed by our constitutional values’. The judgment is thus a morally rich assessment of whether protecting a non-adulterous spouse’s dignity and the sanctity of marriage on one hand, or people’s freedom to choose whom to have sex with on the other hand, is more closely aligned with constitutional values.

A further case, *Gaertner v Minister of Finance*, illustrates more sharply the interplay between strict-sense proportionality analysis and the inquiry into less restrictive alternatives. The case concerned provisions in the Customs and Excise Act 91 of 1964 allowing customs inspectors to enter a wide range of premises without a warrant of court. Key to the Court’s finding that these provisions imposed an unjustifiable limitation of the right to privacy was the conclusion that requiring a warrant for entry into premises where the expectation of privacy is high, such as a person’s home, would be a less restrictive means to enforcing customs and duties than searches without a warrant.

This finding does not stand on its own, however. Some searches, the Court remarked, ‘are generally more invasive and involve a greater limitation of the right to privacy’. The warrantless search of a person’s home is a more serious infringement than the warrantless search of a person’s place of business or warehouse. These more invasive searches in turn require more compelling justifications: while the objectives of the Customs and Excise Act are important enough to justify the less extensive privacy violation of a warrantless search of a business, they are not important enough to justify the warrantless search of a person’s home.

The availability of a less restrictive way of enforcing customs and duties – ie, search on warrant of a person’s home – is not, however, the reason for the Court’s conclusion that the limitation was unjustifiable. Rather, it is a solution to the underlying disproportionality between the serious limitation imposed by warrantless home searches and the importance of enforcing customs and duties. Regardless of whether or not less restrictive alternatives are available, warrantless home searches cannot be justified by the objectives of the Customs

34 *DE v RH* (note 32 above) at para 26.
35 Ibid at paras 54–58.
36 Ibid at paras 62–63.
37 Ibid at para 51.
38 *Gaertner & Others v Minister of Finance & Others* [2013] ZACC 38, 2014 (1) SA 442 (CC).
39 Ibid at paras 68–73.
40 Ibid at para 65.
41 Ibid at paras 62–63.
and Excise Act. It is not the factual availability of less restrictive alternatives that renders the impugned provisions unjustifiable, but the conclusion that constitutional values in this case are better served by prohibiting warrantless searches of homes than allowing them. The Court’s determination that there are less restrictive means to achieve the Act’s objectives is consistent with and supportive of the primary conclusion that the rights limitation is disproportionate, but is not necessary to that primary conclusion.

C  The rule-of-law objection to moral reasoning

The cases I identify in the previous section are examples of a kind of moral reasoning in the Court’s limitations jurisprudence. In these cases, the decision whether to uphold rights or allow their limitation depends on the persuasiveness of arguments that one course of action produces results that are more consistent with the values on which the constitutional order depends and which section 36(1) explicitly articulates.\[1\] In balancing rights and their limitations, section 36(1) instructs judges to consider which option more faithfully advances the Constitution’s commitment to openness, democracy, equality, freedom and dignity.\[2\]

But this form of reasoning faces the criticism that it is unintelligible to anyone but the judge, undermines the certainty of law, and in turn dulls law’s efficacy as a tool for ordering the behaviour of its subjects.\[3\] The premise of this criticism is that if the law is to guide social

\[1\] J Raz ‘On the Authority and Interpretation of Constitutions: Some Preliminaries’ in his Between Authority and Interpretation (2009) 323, 325 and 328. Joseph Raz takes it as a defining feature of constitutionalism that any given constitution reflects ‘principles of government … that are generally held to express the common beliefs of the population about the way their society should be governed’. He argues further, in a passage that evokes some of what I argue is going in the Constitutional Court’s strict-sense proportionality analyses, that constitutional theory purports to explain ‘under what conditions the constitution of a country is legitimate, thus fixing the condition under which citizens have a duty to obey it. In doing that, it provides an account of the principles of political morality that underpin the constitution, in that they justify and legitimize its enforcement. This affinity between moral reasoning and constitutional adjudication in Raz’s theory is the core of the argument I set out in part V, that even less restrictive means analysis depends on moral reasoning.

\[2\] HCJ 14/86 Laor v Israel Film and Theatre Council [1987] IsrSC 41(1) 421 (The Supreme Court of Israel has described the proportionality analysis involved in the justification of rights limitations as an inquiry into whether upholding a right or allowing its limitation is more ‘socially important’.) See also A Barak Proportionality: Constitutional Rights and Their Limitation (2012). Former Chief Justice of Israel Aharon Barak notes in his extra-curial writing that what is socially important, in turn, depends on ‘political and economic ideologies, from the unique history of each country, from the structure of the political system and from different social values’. Ibid at 349. The social importance of protecting a right against limitation or allowing the limitation of a right in a particular case, Barak goes on, is derived from the constitution’s purposes and the degree to which upholding either option ‘advance[s] the legal system’s most fundamental values.’ Ibid at 361.

conduct and provide a framework of rules within which people can act, then those rules need to have certain characteristics. Lon Fuller offers one laundry list of these characteristics, which he famously called the principles of legality: laws must be clear, public, applicable to everyone generally, consistent with one another, non-retroactive, stable over time, and they cannot command the impossible. In addition, public officials’ conduct must be congruent with previously declared rules.45

One of the most basic commitments of the rule of law is that authority over people be exercised only through law, and in accordance with law’s previously promulgated rules. It is the law that should instruct people how to behave, rather than princes, dukes, or highwaymen with guns. And when public officials instruct members of society what and what not to do, those instructions must themselves remain within the limits of the law. If the bureaucratic or judicial administration of a legal system is not sufficiently consistent with Fuller’s principles of legality, the argument goes, it sacrifices this most basic value of the rule of law. Laws that are not clear or public, for example, make it difficult for people to know how they are supposed to behave and makes official enforcement of those laws against them arbitrary and unpredictable. Even if laws meet the first seven of Fuller’s principles of legality, the rule of law will be replaced by the rule of persons if officials’ conduct is incongruent with previously declared rules.

These rule-of-law commitments are just as much an issue in the courtroom as in law-making or the bureaucratic functions of the administrative state. As legal officials, judges are just as much constrained by previously declared rules as any other official. But just about every legal dispute between private parties involves a disagreement about what the law actually allows or requires. In the hard cases where the law has no answer as to how to resolve a dispute, judges must exercise a degree of discretion to decide what the law does mean.46 Although there is an area of indeterminacy at the penumbra of law’s settled core,47 judging in this marginal penumbra can remain consistent with the rule of law if, in Blackstonian fashion, judges make new rules that both settle the dispute before them and apply generally and clearly to similar disputes in the future.48

Judicial decisions that rely on proportionality in the strict sense, however, appear to the critics to evade formulation as general rules applicable to future cases. There is no clear, public, or intelligible standard or metric by which judges determine the moral weight of rights and the

---


competing objectives that rights limitations serve. On the contrary, this rule-of-law objection continues, the weights that judges attach to these competing interests are products of nothing other than the subjective moral intuitions of each judge. The balancing inquiry, as the Canadian constitutional scholar David Beatty puts it, is a ‘freewheeling’ and ‘unprincipled’ moral frolic that neither relies on nor generates clear, stable, predictable, or publicly knowable rules of conduct, and is for that reason incompatible with a commitment to the rule of law.

Although it is not my main concern here, I believe much of this criticism can be allayed by conceiving of strict-sense proportionality analysis not as the direct weighing up or balancing of the value of rights and the value of competing objectives, but as the assessment of each against the constitutional values each system happens to commit to. In any case, the point for present purposes is that these critics of strict-sense proportionality analysis prefer less restrictive means analysis because, on their view, it is more intelligible and better upholds a commitment to the rule of law. It is to this mode of proportionality reasoning that I now turn.

III CHEVRON V WILSON

A The path of least restriction

Not unlike Lawyers for Human Rights, the Court’s reasoning in Chevron v Wilson is not likely to make it into a handbook of constitutional law. This is not because it is faulty or suspect, but rather that the judgment does not add much to the existing jurisprudence. As in Lawyers for Human Rights, the Court mostly treads a path of reasoning already laid down. Unlike Lawyers for Human Rights, however, the conclusion in Chevron rests most apparently on the factual finding that there are means less restrictive than those chosen by the government to achieve the objectives it set out to achieve.

The case involved an arrangement between the operator of a transport company (Wilson) and the supplier of petroleum products to Caltex petrol stations (Chevron) in terms of which the transport company filled its vehicles with petrol at Caltex stations on credit and paid in bulk at the end of each month. At some point a dispute arose as to the accuracy of Chevron’s billing, and after Wilson refused to pay the amounts that Chevron claimed was owing to it, Chevron brought suit in the magistrate’s court. During proceedings it became clear that Chevron was

49 The complaint that strict-sense proportionality analysis involves the comparison or weighing up of incommensurable goods – the ‘incommensurability objection’ – is just one element of the broader concern that strict-sense proportionality analysis, at least when conceived of as balancing, is either inconsistent with the rule of law or simply incoherent. Bendix Autolite Corp v Midwesco Enter Inc 486 US 888, 897 (1988)(Scalia J); N Petersen ‘How to Compare the Length of Lines to the Weight of Stones: Balancing and the Resolution of Value Conflicts in Constitutional Law’ (2013) 14 German Law Journal 1387; S Woolman & H Botha ‘Limitation’ in S Woolman & M Bishop (eds) Constitutional Law of South Africa (2nd Ed, 2008) Chapter 34, at 69–100 (On balancing as an inapt metaphor and the inevitability of incommensurability given the goods and values that must be squared).

50 In addition to the critics listed in note 44 above, a summary of this line of criticism can be found in N Carmi ‘The Nationality and Entry into Israel Case before the Supreme Court of Israel’ (2007) 22 Israel Studies Forum 26, 33; K Möller ‘Proportionality: Challenging the Critics’ (2012) 10 International Journal of Constitutional Law 709, 727–730. Similarly, Ronald Dworkin characterises the objection to discretionary judicial reasoning in hard cases as rooted in the sense that it ‘seems unfair, contrary to democracy, and offensive to the rule of law’. Dworkin Taking Rights Seriously (note 46 above) at 123.


not a registered credit provider in terms of s 40 of the National Credit Act 34 of 2005. The Act provides that credit arrangements made by unregistered credit providers are unlawful from the moment they are concluded, and that any amounts paid to the credit provider under the terms of the agreement must be paid back to the consumer (s 89(5)(a) and (b)). Applying these provisions, the magistrate’s court found itself obliged to order Chevron to pay back to Wilson all the money already paid to it – some R33 million (about US $2.5 million).

Chevron approached the Cape Town High Court and in turn the Constitutional Court for an order declaring that s 89(5) of the National Credit Act amounts to an unjustifiable limitation of the right not to be arbitrarily deprived of property (s 25(1) of the Constitution). Both courts agreed that requiring the refund of monies already paid by a credit consumer by an unregistered credit provider amounts to a deprivation of property, and further that in leaving courts no discretion to make an order that is just and equitable in the circumstances, the deprivation was arbitrary.53

Turning to whether the limitation of the right could be justified, the Constitutional Court began with a statement that genuflects at the altar of proportionality in the strict sense: ‘To determine whether there is sufficient reason for a deprivation’ the Court said, ‘it is necessary to evaluate the relationship between the purpose of the law and the deprivation caused by that law.’54 In the very next paragraph, however, after noting that the purpose of s 89(5) of the National Credit Act is to protect consumers from predatory lenders, the Court fell back on the inquiry into whether there were less restrictive means to achieve that purpose.55 Recognising that a judicial discretion to make an order that is just and equitable in the circumstances is less restrictive than the peremptory injunction to order the full refund of all monies paid, the Court concluded that the limitation of property rights was not justifiable.

The judgment in Chevron v Wilson does not engage in any outright reasoning about or balancing of the moral value of protecting consumers from predatory lenders on one hand, and the moral value of preserving property from arbitrary deprivation on the other. As I argue in part V, I think there is necessarily some moral reasoning at work in cases like this, first of all in concluding that the impugned limitation is disproportionate whether or not there are less restrictive alternatives, and second in assessing whether the less restrictive alternative – in this case a judicial discretion to craft an appropriate remedy – is itself a justifiable limitation of property rights and thus actually a meaningful alternative. But the point for present purposes is that whether or not the less restrictive alternative imposed a justifiable limitation, the Court held that because in fact there was a less restrictive alternative available, s 89(5)(b) imposed a necessarily disproportionate limitation.

Unlike Lawyers for Human Rights where the inquiry into strict proportionality did all the work in finding the limitation unjustifiable, or Gaertner where the inquiry into less restrictive means was tied up with an inquiry into proportionality in the strict sense, the Court here does not even reach the stage of strict proportionality balancing. There are other examples where the Court relies on similar reasoning, but I mention only one here by way of example – the earliest I could find in the Court’s record. In 1995, the Court considered in S v Ntuli whether statutory restrictions on self-represented appeals against criminal convictions by persons already serving prison sentences for those convictions unjustifiably limited rights to fair trial and

53 Chevron v Wilson (note 9 above) at paras 17–24.
54 Ibid at para 32.
55 Ibid at para 33.
equality before the law.\textsuperscript{56} Section 305 of the Criminal Procedure Act 51 of 1977 provided that self-represented appeals could proceed only if a high court judge certified that there were reasonable grounds for appeal.\textsuperscript{57}

Having found that the provisions did indeed limit fair trial and equality rights, the Court then considered if the limitations were justifiable. Identifying the purpose of the provisions as avoiding the crowding of courts rolls with frivolous appeals at the expense of meritorious ones, the Court concluded that the requirement for a judge’s certificate was more restrictive than it needed to be because it presumes, without evidence, that only the frivolous appeals will get stopped. Since a variety of alternative mechanisms existed for controlling frivolous appeals,\textsuperscript{58} the Court found that the mere possibility that the need for a judge’s certificate would stop meritorious appeals was enough for it to conclude that ‘the means used to achieve the end therefore go beyond it’ and were disproportionate.\textsuperscript{59}

At no point, however, did the Court consider whether the objective of reducing the number of frivolous applications for appeal was proportional, in the strict sense, to the limitation on rights that the requirement for a judge’s certificate imposed. That is, the Court did not ask whether the extent of the limitation on the right to appeal imposed by the requirement of a judge’s certificate was justified by the benefits of a less crowded court roll.

B De-moralising proportionality analysis

One way of looking at cases like \textit{Ntuli} and \textit{Chevron v Wilson} is that, sometimes, a court need not reach the morally rich inquiry into proportionality in the strict sense in order to find that the limitation is unjustifiable. If a rights-limiting measure is unlikely even to achieve its purpose, then surely there can be no justification at all for the limitation. Similarly, if the same objectives that a rights-limiting measure aims at can be achieved in a manner that is less restrictive of rights, then surely the less restrictive approach is preferable and the more extensive limitation is not justified.

Fact-centred reasoning of this kind eliminates – or purports to eliminate – the moral reasoning that strict-sense proportionality analysis involves. To the extent that less restrictive means analysis involves no moral judgments, it is able to sidestep the rule-of-law objection levelled against strict-sense proportionality analysis. The core of this rule-of-law objection is that judges’ subjective moral reasoning suffers from the same unintelligibility that for Jeremy Bentham runs through the common law as a whole.\textsuperscript{60} The moral intuitions of judges are not accessible to or intelligible to the subjects of judicial decisions, and in many cases not necessarily reflective of or shared by the members of the society subject to those decisions. Allowing judges to make decisions on the basis of these inaccessible intuitions, so the objection goes, undermines the authority of law because those decisions are not persuasive to those who do not share the same moral intuitions.

If hard cases about rights limitations can be solved without the impressionistic, subjective, ad hoc and unprincipled moral reasoning that strict-sense proportionality analysis requires judges to engage in, and rather according to rules that are more transparent and intelligible

\textsuperscript{56} \textit{S v Ntuli} [1995] ZACC 14, 1996 (1) SA 1207 (CC).
\textsuperscript{57} Ibid at para 4.
\textsuperscript{58} Ibid at para 28.
\textsuperscript{59} Ibid at para 24.
\textsuperscript{60} See J Bentham \textit{A Fragment on Government} (note 49 above).
than judges’ moral intuitions, then the commitment to the rule of law and the principles of legality will be better upheld.\(^6^1\) Judicial decisions can be more easily justified to those they affect, the argument goes, if their conclusions depend on objectively verifiable evidence and facts about the world.\(^6^2\)

At the same time, the fact-focused decision rules of the less restrictive means inquiry constrain whatever discretion judges do have to make new law in the penumbra of legal uncertainty. Judicial decisions about limitations guided by what the evidence suggests is the least restrictive approach to achieving some set of objectives, are less influenced by whatever moral commitments a judge happens to have. And precisely because less restrictive means analysis decentres moral reasoning from limitations analysis, the argument goes that court judgments are more intelligible and more easily justified to legal subjects when they rely on less restrictive means analysis to the exclusion of strict-sense proportionality analysis. It is precisely this de-moralising exercise – the removal of moral reasoning from proportionality analysis – that purports to inoculate it against the rule-of-law objection. On this argument, a judgment like *Chevron v Wilson* is more intelligible, more easily justified to legal subjects, and thus more consistent with the rule of law than a judgment like *Lawyers for Human Rights*.

**IV PROPORTIONALITY ANALYSIS AND THE CONCEPTION OF LAW**

The South African Constitutional Court’s record reflects how it has relied on both less restrictive means analysis and strict-sense proportionality analysis in considering the constitutionality of rights limitations.\(^6^3\) A court’s reliance on less restrictive means analysis in some cases and strict-sense proportionality analysis in others can be explained, I think, by the court’s view as to which mode of reasoning is more intelligible to legal subjects in each case, and thus which produces the more persuasive judgment.

If we accept the idea that a judicial decision must be intelligible to legal subjects if it is to be an authoritative statement of law, then this debate about the relative intelligibility of modes of proportionality analysis is important. The fact that there is an ongoing debate about whether one mode of reasoning is more intelligible, more predictable, more consistent with the rule of law and so on, suggests that we are indeed committed to the idea that judicial decisions must be intelligibly justified.\(^6^4\)

\(^6^1\) D Beatty *The Ultimate Rule of Law* (note 51 above) at 74.


\(^6^3\) N Petersen notes that the Court has tended to favour a more fact-based reasoning of less restrictive means analysis in order to strike down rights limitations contained in legislation. *Proportionality and Judicial Activism* (note 7 above) at 109–110. In contrast, the Court has been more willing to rely on strict-sense proportionality analysis to strike down provisions in the common law and to uphold as justifiable statutory rights limitations.

I am not interested at this point, however, in which form of reasoning is in fact more intelligible, or even if measuring such a thing is possible at all. Rather, I consider in this part of the article how each of these paths of proportionality analysis makes a claim to intelligibility, and in turn to authority, and how those claims connect each path of proportionality analysis to different conceptions of law and of the rule of law. In part V which follows, I argue that the differences between these conceptions are actually not all that stark, and that less restrictive means analysis ultimately relies on much the same kind of moral reasoning as proportionality in the strict sense. For both modes of reasoning, intelligibility and authority depend on making arguments that are morally persuasive, rather than merely factual persuasive.

A Raz’s positivism: a service conception of the constitution

Consider again how less restrictive means analysis claims to meet the rule-of-law objection by replacing substantive moral reasoning about the value of rights and competing limitations with a factual calculus about which course of action impairs rights less. In doing so, less restrictive means analysis fits the mould of what Joseph Raz calls the ‘service conception of authority’. Raz sees the service function of the law as providing rules for conduct which are binding on us, or authoritative, because they claim to align our conduct more closely to the legal system’s basic normative commitments than do our own judgments about how we ought to behave. We defer to legal authorities for guidance on how we should behave – courts in cases of disputes, legislatures in the case of our everyday behaviour – in order to avoid having to engage for ourselves in complicated moral reasoning about how we ought to act. ‘The whole point and purpose of authorities’, Raz says, ‘is to preempt individual judgment on the merits, and this will not be achieved if in order to establish whether the authoritative determination is binding individuals have to rely on their own judgment of the merits.’ Raz calls this the ‘pre-emption thesis’.

Along with the pre-emption thesis, Raz’s service conception of authority combines two other ideas: the ‘dependence thesis’ and the ‘normal justification thesis’. The dependence thesis captures the idea that there are already reasons for people to act in particular ways, and on which law’s normative force depends. These reasons for action would apply to people in the absence of the law, and legal directives or rules should therefore summarise and reflect these already existing reasons. The ‘normal justification thesis’ maintains that law is justified, and thereby claims authority, because people are likely to better comply with these already existing reasons for action if we follow the law’s directives than if we try to work out for ourselves what those reasons or norms require us to do. The law generally and legal rules and judgments specifically are meant to replace or pre-empt the reasons on which they depend, giving the subjects of the law a new reason for acting in one way or another.

In Raz’s view it is easier, for example, to pay the taxes that the tax agency tells us to pay than to try and work out for ourselves what percentage of our earnings would be just or reasonable to redistribute or to give to the civic institutions that provide services like sewerage, waste removal, health care and social security. The service of law is that it absolves ordinary people from having to deduce from first principles how we should behave. Instead of engaging in primary moral reasoning ourselves, we can rely on legal authorities to tell us what to do because

65 J Raz ‘Authority and Justification’ (note 11 above) at 15.

Constitutional Court Review 2019

235
all the moral calculations have already been done, by the legislature in formulating the law or by the courts in applying the law.

It is important to note that for Raz, however, the point is not that law must in fact do a better job than we could do ourselves of aligning our conduct with norms, but that a legal system must necessarily make the claim that it does so. Moreover, a legal system must display the qualities that make such a claim possible. ‘To claim authority’, Raz argues, a legal system ‘must be a system of a kind which is capable in principle of possessing the requisite moral properties of authority.’ A central feature of such a legal system, he goes on, is an identifiable and intelligible set of legal rules. Law is normally justified – that is, acquires ‘legitimate’ rather than merely ‘de facto’ authority – by its capacity to intelligibly convey to us how already existing norms require us to behave, and to pre-empt those norms as reasons for action. To this end, law must in the first place be identifiable. Of course, lawyers, judges and legal subjects will continue to engage in some degree of argument about how the law applies to particular disputes and how the law might require us to act in particular circumstances, but the answers to those inquiries no longer depend, for Raz, on moral reasoning about the normative commitments that happen to underlie the law.

This leads Raz to a defence of the ‘sources thesis’, a central theme in the tradition of modern legal positivism. The sources thesis posits that the only determinant of a legal rule’s validity is its source. Legal validity is on this view a matter entirely separate from the question of law’s moral virtue, and the latter simply plays no role in determining if law is valid or not. By looking only to a purported legal rule’s provenance, we can more certainly identify what is properly considered law that we have reason to obey, than if we were to engage in arguments from first principles about the moral virtues of purported legal rules. As subjects of the law, Raz insists, we need do no more than follow the rules promulgated by recognised sources of law, and we have no business calling into question on moral grounds the authority of laws promulgated by recognised sources. Legal positivism strives for a legal system that is a ‘settled, public and dependable set of standards for private and official conduct’. By eliminating moral reasoning from the assessment of a rule’s legal validity, the legal system gains the certainty and clarity Raz says it needs to make the claim to authority.

For a court judgment about the constitutionality of a rights limitation to claim authority, on this Razian view, it must provide an intelligible reason for its conclusion that pre-empts any reasoning from moral first principles. Questions about the constitutionality of rights limitations thus elevate the consideration of authority to the constitutional level: while Raz is concerned with how ordinary statutes and court decisions authoritatively mediate between

67 Ibid at 300.
68 Ibid at 296–297.
69 Ibid at 304–306.
71 R Dworkin Taking Rights Seriously (note 46 above) at 347. It might look like something of a scarecrow argument to rely on one of legal positivism’s greatest critics for a description of it, but Raz himself admits that Dworkin’s description of this central element of legal positivism is exactly the account that he wants to defend. J Raz ‘Authority, Law and Morality’ (note 66 above) at 320.
people and reasons for action, a challenge to the constitutionality of a statute concerns how the constitution and constitutional adjudication authoritatively mediate between the legislature and fundamental constitutional principles. In the same way that normally justified law purports to provide pre-emptive answers to questions about how we ought to behave, replacing any primary moral reasoning we might do in working out what the law actually is, a proportionality analysis that asks only whether there are in fact less restrictive means to achieve an important objective purports to provide pre-emptive answers about the justifiability of rights limitations, absolving courts from having to do any moral reasoning as to whether upholding a right or allowing its limitation better advances the constitution’s normative commitments.

Instead of wading into moral debates about the relative value of rights and the policy objectives a limitation aims to achieve, or about the relative harm of limiting a right or choosing not to pursue an incidentally rights-limiting policy objective, courts need only apply a mechanical-decision rule to the facts of the case and report the results. Less restrictive means analysis claims authority for constitutional adjudication to the extent that it produces comprehensible and intelligible reasons for its conclusions. It is the best kind of sausage-making: judges simply feed the factual ingredients into a machine and collect the decisional sausage at the other end, without ever having to think about how sausages are actually made inside the machine or get their hands dirty with whatever sausages are made of. In this service conception of the constitution, the outcomes of fact-based reasoning pre-empt consideration of moral first principles, which judges would otherwise have to try to make sense of in order to decide whether to uphold a right or allow its limitation. And given enduring and reasonable disagreement about the content of our moral commitments, Raz sees this pre-emption of primary moral reasoning as a service indeed.

For Raz, law claims authority because it guides us both to do the things we have reasons to do anyway, and to comply more fully with those reasons than if we were left to work it out for ourselves. The law, in other words, is more intelligible to us as a reason for action than its underlying moral imperatives. Judgments based on less restrictive means analysis, for their part, claim authority to the extent that they uphold underlying constitutional commitments more intelligibly than decisions purporting to explain in moral terms what those constitutional commitments require. The service conception of the constitution thus maps onto Raz’s more general conception of the law as a set of authoritative rules that require no moral reasoning to identify.

The Razian flavour to a court’s preference for less restrictive means analysis, then, is that the subjective, impressionistic and allegedly less intelligible moral reasoning inherent in strict-sense proportionality analysis undermines the conditions necessary for judgments on rights limitations to claim authority. While fact-centred decision rules allow judges and legal subjects alike to readily identify what the constitution requires, pre-empting moral arguments about what our basic constitutional commitments mean, strict-sense proportionality analysis seems to require exactly these moral arguments. The pre-emptive force of strict-sense proportionality analysis is low, for its part, because it turns on precisely those fundamental questions of morality that Raz’s service conception aims to pre-empt. A legal system full of strict-sense proportionality analysis thus risks sacrificing the conditions necessary to make a claim to authority. On this Razian conception of law, rights adjudication can make a claim to authority only by minimising the room for moral reasoning.

72 Raz ‘On the Authority and Interpretation of Constitutions’ (note 42 above) at 369–370.
B Fuller’s morality: justifying limitations against constitutional commitments

It is the moral reasoning inherent in strict-sense proportionality analysis that makes it a target of the rule-of-law objection. But this objection, I think, rests on an impoverished conception of the rule of law that does not recognise that legal subjects’ moral agency is at the very foundation of the commitment to the rule of law.73 Strict-sense proportionality analysis aligns with a conception of law that is closely connected to Fuller’s ideas about the rule of law, rather than the positivist conception of law with which less restrictive means analysis aligns. The key to distinguishing between these different conceptions of law is the way that each understands the rule of law. The problem for proponents of less restrictive means analysis, however, is that Raz endorses a conception of the rule of law that is ultimately impossible to separate from Fuller’s. As I argue further, this affinity between Raz and Fuller supports a further affinity between the two paths of proportionality analysis, in that both ultimately involve unavoidably moral reasoning.

I start with the question of why we care about the rule of law in the first place. As subjects of the law, we value clarity and intelligibility in the operation of law not simply because these are good things in themselves for a legal system to have, but because a clear system of rules allows us to make meaningful decisions about our lives. An unintelligible legal system negates the value of people’s choices because it makes it more likely that our choices will be frustrated by some legal rule or decision we did not understand or whose implications we could not have anticipated. The instrumental value of a clear and intelligible legal system is that it guarantees that the law respects our choices.

Raz’s conception of the rule of law follows this instrumental account. The rule of law is not something that a system of rules must possess in order to qualify as a legal system. On the contrary, Raz sees the rule of law as just one of the stars in the constellation of legal ideals.74 It is a valuable quality for a legal system to have, in part, because it makes law more effective as a technology of government. Just as sharpness is the ‘specific excellence’ of knives, the rule of law is the specific excellence of law because the more a legal system sticks to rule-of-law requirements like clarity and intelligibility, the more effective it will be in communicating to us what is allowed and what is proscribed.75

Raz also agrees with Fuller that adherence to the rule of law curbs arbitrary power. Raz describes the rule of law as a negative virtue because it limits the danger of arbitrary power that the law enables in the first place,76 while Fuller contends that adherence to the rule of law limits opportunities for public evil because a conscientiously constructed and administered legal system ‘exposes to public scrutiny the rules by which it acts.’77 Raz accepts, then, that the rule of law is valuable because it constrains the operation of law to stable and predictable channels, minimising unexpected exercises of power and creating room for the meaningful and

74 Raz ‘The Rule of Law and its Virtue’ (note 45 above) at 211; J Waldron ‘The Classical Lockean Picture and its Difficulties’ Lecture 1 in The Rule of Law and the Measure of Property (note 44 above) at 1, 12.
75 Raz ‘The Rule of Law and its Virtue’ (note 45 above) at 225.
76 Ibid at 219–220.
77 Fuller The Morality of Law (note 10 above) at 158. See also I. Fuller ‘Positivism and Fidelity to Law: A Reply to Professor Hart’ (1958) 71 Harvard Law Review 630, 636.
productive exercise of individual agency. The rule of law has value, even for Raz, because it affirms a view of human beings as morally autonomous agents capable of rational action. He warns in this light that to depart from the rule of law is to undermine human dignity by disrespecting legal subjects’ moral autonomy.

The difference between Fuller and Raz is that Fuller sees the rule of law not as a virtue to be sought pragmatically, but a quality without which a legal system cannot operate at all. Law as such is internally moral, in Fuller’s most famous contribution to legal theory, because it is ‘constitutively dependent’ on the view of legal subjects as morally autonomous agents. While for Raz it is a contingent good for a legal system to affirm this view of humanity, for Fuller the law can fulfill its function – that is, it can succeed in shaping social behaviour – only if it presupposes that as legal subjects we have the capacity to understand rules and conform our behaviour to them. Law, as such, affirms individual moral autonomy, conceiving of every legal subject as ‘a responsible agent capable of understanding and following rules.’

As Kristen Rundle puts Fuller’s position:

For Fuller there can be no meaningful concept of law that does not include a meaningful limitation of the lawgiver’s power in favour of the agency of the legal subject. This is not a moral objective that is imposed in the enterprise of lawgiving from without it. It is, rather, simply something that follows from the formal distinctiveness of law as the enterprise of subjecting human conduct to the governance of general rules.

From the perspective of constitutional adjudication, the implications of this view of legal subjects are far-reaching. If we find predictability and intelligibility in the legal system valuable because of its service to moral autonomy and the capacity for reason in which individual choice is rooted, then it must be the case that autonomy and the capacity to reason are the primary goods to which predictability and intelligibility are secondary. An intelligible legal system, and in turn intelligible legal reasoning, is valuable only to the extent that it tends to facilitate the exercise of moral autonomy.

Accordingly, law must meet the demands that morally autonomous, rational, and free-thinking people put on it. For Fuller, this requires that legal rules display the characteristics he articulates in his eight principles of legality. Our capacity to act as morally autonomous agents who formulate and pursue life plans just as Raz thinks we do, is seriously undermined when the law changes from day to day, is impossible to understand or to comply with, applies retroactively or is kept secret, or if official behaviour is incongruent with previously declared rules.

But Fuller’s principle of congruence encompasses both the highly specific and formal rules of law that make up statutes and regulations, and the fundamental normative commitments on which a legal system rests. Where a constitution happens to make a commitment to a set of fundamental normative values (openness, democracy, dignity, equality and freedom, for example), the principle of congruence demands that official conduct remain congruent with these values as well as with formal rules. Every failure of congruence, whether a departure

---

78 Raz ‘The Rule of Law and its Virtue’ (note 45 above) at 220–222. See also K Rundle ‘Form and Agency’ (note 12 above) at 776.
79 Raz ‘The Rule of Law and its Virtue’ (note 45 above) at 222.
80 Rundle ‘Form and Agency’ (note 12 above) at 771.
82 K Rundle ‘Form and Agency’ (note 12 above) at 772.
from the formalities of duly promulgated statutes and regulations or from fundamental constitutional values, is for Fuller and for Raz an affront to legal subjects’ moral autonomy because it undermines the settled order within which we make decisions about what to make of our lives.

To meet the rule-of-law demand for normative congruence, a government must, in the first place, act in ways that tend to uphold the values to which the constitution happens to commit a political community. But more than this, the commitment to the view of legal subjects as morally autonomous agents requires the government to demonstrate to us, through a process of persuasion and argumentation, that its actions – rights-limiting legislation, for example – are congruent with constitutional values. For government to act only in ways that are demonstrably congruent with underlying values and previously declared rules supplies normative reasons for legal subjects to accept the authority of the law, in turn generating a ‘bond of reciprocity’ between lawgivers and legal subjects.83 Strict-sense proportionality analysis is a mode of argumentation tailored to demonstrating whether limiting a right in pursuit of some objective or striking down the limitation and upholding the right better advances the values set out in the constitution.

The value of the rule of law is not only the clear, intelligible and predictable legal system it may generate, as a matter of fact. We value the rule of law and demand normative congruence because, beyond these qualities, it requires the legal system’s authorities to demonstrate to us on the basis of rational explanation and argument that laws are oriented toward the fulfilment of the most basic normative commitments of the political community of which we form part. The view of human beings as morally autonomous rational agents who demand justification for the rules that bind us informs a conception of law which, unlike the service conception of the constitution and the legal positivist tradition in which it is steeped, emphasises its moral and normative content.

Strict-sense proportionality analysis responds to the rule-of-law demand that power be exercised according to law, where ‘law’ as such is understood as having to be justifiable against a legal community’s most deeply held moral convictions. ‘Legality can produce legitimacy’, Habermas suggests, ‘only to the extent that the legal order reflexively responds to the need for justification that originates from the positivization of law and responds in such a manner that legal discourses are institutionalized in ways made pervious to moral argumentation’.84 Strict-sense proportionality analysis is a tool that allows courts and policy makers to make deeply moral arguments explaining decisions with reference to the values that a constitution entrenches at the foundation of the legal order, and thereby to claim authority for their decisions.

V THE MORALITY OF LESS RESTRICTIVE MEANS ANALYSIS

I argued in part IV above that each of the two paths of proportionality analysis that the South African Constitutional Court has travelled since 1994 corresponds to a particular understanding of how law in general, and legal judgments in particular, make a claim to authority. Both modes of reasoning purport to make limitations analysis intelligible to the

83 L Fuller Morality of Law (note 10 above) at 39–41 and J Waldron ‘Why Law – Efficacy, Freedom, or Fidelity?’ (1994) 13 Law and Philosophy 259, 275–280 (Jeremy Waldron describes the valuable core of the rule of law as the ‘fidelity to law’ that it generates.)

subjects of the law: less restrictive means analysis in Razian fashion by replacing primary moral reasoning with factual analysis, and strict-sense proportionality analysis in Fullerian fashion by arguing that either upholding or striking down a rights limitation is more congruent with basic constitutional values.

Despite these different approaches to law’s authority, Raz nevertheless understands the rule of law, as Fuller does, to presuppose and affirm legal subjects as morally autonomous agents capable of rational thought and conduct. This calls into doubt the extent to which less restrictive means analysis, aligned as it is with Raz’s service conception of authority, is ultimately capable of pre-empting judges’ moral reasoning in limitations cases.

The core of the case I make here is that while less restrictive means analysis may operate on facially factual grounds, it can do so only because the moral reasoning has already been done outside the purview of the less restrictive means inquiry. Less restrictive means analysis does not pre-empt moral reasoning after all, so much as proceed as if the moral conclusions are foregone. Upholding a rights limitation because it has adopted the least restrictive means presupposes that the objective the limitation purports to achieve is important or valuable enough to justify limiting a right (or at least, to justify limiting it minimally). But the less restrictive means analysis does not by itself help us to understand why limiting a right even to that minimal extent is justified by the objective the limitation hopes to achieve: the moral conclusion that the minimally restrictive limitation is justified – i.e. that it is proportionate to the harms it causes – must have already been reached by analytical means other than the less restrictive means inquiry itself.85

Indeed, in order to conclude that the government’s chosen means of achieving some objective is disproportionate because there is some less restrictive alternative available, it must be the case that that less restrictive alternative is itself proportionate. If it were not, it would not be an alternative open to the government. The difficulty for less restrictive means analysis, however, is that this conclusion about the proportionality of the less restrictive alternative cannot depend on less restrictive means analysis. That would be terribly question-begging. Even where less restrictive means analysis seems to do all the work, in indicating as a matter of fact that there are other ways to achieve an objective that do less damage to rights, the foundations of that analysis are laid by a different mode of reasoning – namely, strict-sense proportionality analysis.

A similar gap in the logic of less restrictive means analysis appears in cases where the adoption of the least restrictive means fails to resolve moral disagreement, or at least fails to persuade legal subjects to put aside whatever moral objections they may have and accept the authority of the decision. Former Israeli Supreme Court Justice Aharon Barak provides an example which corresponds roughly to the situation the South African Constitutional Court confronted in the S v Walters reference. If shooting dead a fleeing criminal is indeed the only, and thus least restrictive way to apprehend a criminal, then the reliance on the less restrictive means test alone compels us to accept it as a justifiable rights limitation.86


Whether using deadly force to stop a fleeing criminal is justified or justifiable in certain cases is a serious question, however, about which people may reasonably disagree. To the extent that a limitations inquiry ends with the purely factual conclusion that there are no less restrictive means available, it leaves no room for discussion of these moral disagreements. All this indicates, a proponent of less restrictive means analysis might say, is that a court should not uphold a rights limitation unless it has found that it is both minimally impairing and proportionate in the strict sense: it remains open to a court to strike down a limitation only because it does not adopt the least restrictive means. But the argument cannot go both ways. There is something self-defeating, or at least internally inconsistent, in asserting on the one hand that a court must avoid moral reasoning in concluding that a less restrictive route to achieving a government objective is a proportionate and thus acceptable alternative to the government’s chosen means, and yet demanding on the other hand that a court engage in moral reasoning when an imitation is minimally impairing.

Raz for his part acknowledges that ‘legal authority is itself a form of claimed moral authority’, to the extent that it relies on already established moral principles. A great deal of legal argument is consequently ‘technical’ rather than moral, and involves merely explaining what the position in the law actually is. Once the moral justification of a system’s ultimate legal rules is established or assumed, Raz goes on, ‘the moral justification of the rest of the law is – up to a point – established by technical legal argumentation.’ Less restrictive means analysis is persuasive as a form of technical legal argumentation, and makes a meaningful claim to authority, only because moral justifications of the principles on which a particular less restrictive means analysis depends have already been provided elsewhere.

Raz admits that sometimes we do appeal to moral reasons to justify a claim to legal authority, even if we mostly rely on technical argument. Constitutional adjudication presents a special case for Raz, however. Constitutions are morally and legally under-determinative and vague on many issues. In many cases a constitution will provide no clear answers, in the face of people’s inevitable disagreement on questions of morality and the indeterminacy of claims to moral truth. In these situations, Raz concludes, formal legal doctrine that is ‘detached from moral considerations’ should emerge to serve as a ‘distancing device’ between moral disagreement and the intelligible resolution of legal disputes. It seems to me that less restrictive means analysis is one of these distancing devices which for Raz circumvents any need to morally justify a legal decision. But it is difficult to understand how the non-moral and purely technical justification of a decision that engages precisely those moral matters about which people deeply disagree could be shown to be justifiable to them without any moral argumentation at all. On the contrary, the fact of moral disagreement demands moral argumentation rather than its avoidance.

Recall that in Gaertner, a case I discuss above alongside Lawyers for Human Rights, the Court found that warrantless home searches in the enforcement of customs and excise legislation unjustifiably infringed the constitutional right to privacy. Its conclusion was supported by its finding that a rule requiring home searches be conducted only with a warrant of court would be less restrictive of privacy. But finding that there are, in fact, less restrictive alternatives carries

---

87 Raz ‘The Authority of Constitutions’ (note 42 above) at 331.
88 Ibid.
89 Ibid at 365.
90 Ibid at 346.
91 Ibid at 369–370.
the necessary conclusion that the harm flowing from less restrictive, on-warrant home searches of a person’s home is proportionate to the value of enforcing customs and excise legislation. Proportionality in the strict sense does all of the argumentative work here, even though the Court relies on less restrictive means analysis.

The Constitutional Court’s own record indicates that the inquiries into less restrictive means and strict-sense proportionality are conceptually distinct, and that less restrictive means analysis by itself will sometimes deliver no answer or the wrong answer. The Court has found, for example, that limitations can be disproportionate in the strict sense regardless of whether they are minimally impairing,92 and perhaps more pertinently has upheld a rights-limiting measure as proportionate in the strict sense even though less restrictive alternatives were available.93

The point is not just that less restrictive means analysis and strict-sense proportionality analysis are distinct inquiries and may on occasion support different results. Rather, the important observation is that a conclusion about a rights limitation’s constitutionality cannot be grounded entirely in less restrictive means analysis, because there are important questions about the moral relationship between rights and limitations that the less restrictive means analysis leaves unanswered.

And if we understand the rule of law and the notion of law’s authority as both Fuller and Raz do, then the reason that judgments need moral justification beyond merely factual justification becomes apparent. It is only by making arguments that appeal directly to legal subjects’ capacity as autonomous moral agents that judgments can be justified against, and claim authority in light of, a society’s normative commitments. Deploying Raz’s service conception of authority in the context of rights adjudication – what I have called the service conception of the constitution – flies in the face of Raz’s own conception of legal subjects as moral agents who demand the moral justification of law.

This is not to say that the factual conclusions that less restrictive means analysis generates are never persuasive or authoritative, but rather to emphasise that the persuasive force of those factual conclusions presupposes that any deep-seated moral disagreements about a rights limitation – that is, disagreements about its strict-sense proportionality – have already been addressed. For a judgment about a rights limitation to claim authority, then, it must be possible to explain why the outcome is congruent with constitutional values. Strict-sense proportionality analysis provides a mechanism to provide this explanation in unashamedly moral terms. Of course, a court’s account of a limitation’s strict-sense proportionality or disproportionality may not resolve these moral disagreements, but the argument in support of that conclusion gives people a reason to accept it as authoritative even if they disagree with it. Less restrictive means analysis, understood as an exercise in pure facts, is not able to provide these moral justifications. Its claim to authority can only be supported alongside the morally rich mode of reasoning that strict-sense proportionality analysis provides.

92 DE v RH (note 32 above); S v Niemand (note 18 above); In re: S v Walters (note 18 above).
93 S v Dlamini; S v Dladla; S v Joubert; S v Schietekat [1999] ZACC 8, 1999 (4) SA 623 (CC)(With respect to limitations on due process rights in order to provide magistrates with information on which to determine bail conditions, the Court looked to foreign jurisdictions for examples of similar provisions that restrict due process rights less, but nevertheless concluded that limitations in question were proportionate in South Africa).
VI CONCLUSION

The rule-of-law objection to strict-sense proportionality analysis is rooted in claims that the subjective moral reasoning it involves is unintelligible. Less restrictive means analysis, the other path of proportionality reasoning that the Constitutional Court has walked since 1994, promises to avoid this objection because it purports to be a factual and value-free heuristic for adjudicating rights limitations cases. In pre-empting primary moral reasons for upholding or striking down rights limitations, less restrictive means analysis presents a Razian view of the authority of legal judgments. What I call the service conception of the constitution separates legal reasoning from questions of morality, and locates legal judgments’ authority in the factual reasoning that pre-empts judges’ moral reasoning.

But the conception of the rule of law at the heart of this attack on strict-sense proportionality analysis is not the only game in town. Lon Fuller offers a normatively rich account of the rule of law which requires just as much congruence between official conduct and the basic normative commitments a legal system happens to make, as it does between official conduct and previously declared rules. Moreover, law’s authority depends on the justifiability of its operation against these basic normative commitments.

I am not concerned here to defend Fuller’s conception of the rule of law against all comers. Rather, I want to point out that the conception of the rule of law that underlies Fuller’s concept of law, and which in turn aligns with strict-sense proportionality analysis, bears a striking resemblance to the conception of the rule of law that Raz presents. Both Raz and Fuller see the value of the rule of law in the service it does individual moral autonomy. Both Raz and Fuller take legal subjects to be undeniably rational agents, suffused with a capacity to think and reason for themselves. For both Raz and Fuller departing from the requirements of the rule of law undermines human agency, which for Raz is merely unfortunate and undesirable but which for Fuller makes impossible the entire project of governing by law.

On the view of the rule of law that Raz and Fuller share, then, the authority of court judgments about rights limitations depends on morally rich reasoning that speaks to and affirms the capacity of legal subjects to reason about the law, understand it and conform their behaviour to its commands. Constitutional rights are inherently moral legal artifices, closely tied to the values and norms that a constitution happens to articulate and to which it commits the legal system. A court is better placed likely to succeed in justifying a decision that upholding a right or allowing its limitation is more closely aligned with these constitutional commitments, and thereby to make a claim to authority, if it can explain why those values pull in one direction rather than another.

Less restrictive means analysis, at least as its proponents describe it, tries to eliminate precisely these morally rich explanations. The difficulty that it has in doing so, I argue, is that the Razian service conception of the constitution it adopts is inconsistent with Raz’s own view about the rule of law and the morally autonomous nature of legal subjects. I do not want to make a broad attack on Raz’s service conception of authority here, but I do want to suggest that it breaks down at the point where it applies to the justification of rights limitations. The breakdown occurs because relying on the factual inquiries of the less restrictive means analysis alone leaves unexplained why less restrictive approaches are themselves justifiable rights limitations. People want to know why the sausage machine spits out each particular

---

94 For a somewhat broader attack than mine, see Rundle ‘Form and Agency’ (note 12 above).
sausage: we want to be convinced that the decisional sausage is composed of ingredients that are normatively palatable to us. Less restrictive means analysis, and the service conception of the constitution on which it depends, does not clear a path for judgments on rights limitations to make a meaningful claim to authority.

I end by noting that moral reasoning remains important to many of the jurisdictions where the constitution provides explicitly for the limitation of constitutional rights. Even where courts rely formally on the less restrictive means analysis and make no overt attempt to balance the value of rights and the objectives of rights-limiting measures, the reasoning is difficult to make sense of without tracing the implicit moral evaluation going on behind the scenes of less restrictive means analysis.95 Ultimately, it may not even be possible for a method of rights adjudication to fully reflect the Razian version of legal positivism. But if South Africa’s legal system is to continue to promote a culture of justification and to move away from a culture of authority, it seems that an approach to proportionality analysis that emphasises the justification of legal decisions and the operation of legal rules against the fundamental values the Constitution articulates is better than one that relies on factual, evidence-based and value-neutral reasoning.

95 Stacey ‘The Magnetism of Moral Reasoning’ (note 6 above).
Forcing the Court’s Remedial Hand: Non-Compliance as a Catalyst for Remedial Innovation

HELEN TAYLOR

ABSTRACT: This article juxtaposes the pioneering remedies granted by the Constitutional Court in its trilogy of Black Sash judgments with similar innovations forged by our lower courts in the cases of Madzodzo, Linkside and Mwelase. By juxtaposing these cases in this way, their shared remedial predicament is foregrounded: litigation progresses in multiple stages as the court is confronted by the government’s continuing failure to comply with its positive constitutional duties, with remedial escalation culminating in a resort to novel mechanisms that seek to compel compliance. In seeking to capture this dynamic in which persistent non-compliance serves as a catalyst for remedial innovation, the article explores two lines of remedial development. First, it considers the emerging use of court-appointed agents to enhance the court’s supervisory jurisdiction, comparing the appointment of the Auditor-General and Panel of Experts in Black Sash I with the independent auditor in Madzodzo, the claims administrator in Linkside and the special master in Mwelase. Second, the resort to a personal costs order against the Minister in Black Sash III is evaluated as a strategy to ‘pierce the political veil’; and is compared to the attachment of state assets in Linkside. This analysis yields significant insights for our remedial jurisprudence. The front-line remedial experimentation in lower court litigation contextualises the innovations that filter up to the Constitutional Court, thus providing a richer understanding of the cross-fertilisation in remedial development between the different tiers of court. The two lines of remedial development also underscore the range and complexity of reasons for government non-compliance. Judicial responses to persistent non-compliance should not be confined to the individual mental states of public officials, but also target the systemic functioning of institutions that chronically fail to fulfil their constitutional obligations.

KEYWORDS: remedies, supervisory jurisdiction, special master, personal costs, attachment of assets

AUTHOR: DPhil candidate, Faculty of Law, University of Oxford, and Law Clerk of the Constitutional Court of South Africa. Email: helen.taylor0412@gmail.com

ACKNOWLEDGEMENTS: I am grateful for the generous feedback received at various stages of this paper’s development: the encouraging engagement with an early draft by participants at the ninth Constitutional Court Review Conference; the constructive suggestions offered by two anonymous reviewers; and the patient editorial support given by Michael Bishop and Raisa Cachalia in bringing this paper to publication. My argument has also been enriched by input from Sandy Fredman and conversations with Sarah Sephton and Stu Woolman, to whom I extend my particular thanks.
I INTRODUCTION

For any remedies enthusiast, the series of Black Sash judgments handed down by the Constitutional Court during 2017 and 2018 makes for binge-worthy reading.¹ In this sequel to the AllPay series,² the Court dealt with the remedial repercussions of an organ of state’s non-compliance with the terms of its order in AllPay Remedy. Black Sash I delivered a stinging judicial rebuke of the South African Social Security Agency’s (SASSA) failure to comply with its undertaking to the Court that it would be able to pay social grants from 1 April 2017. The Court expressed its frustration at being forced to escalate remedial measures in order to avert the ‘national crisis’ caused by SASSA’s non-compliance.³ Forging ahead along the pioneering pathway of the AllPay Remedy judgment,⁴ the Black Sash trilogy broke new ground in our remedial jurisprudence through its innovative response to persistent non-compliance. In Black Sash I, the Court reinforced its supervisory jurisdiction with the appointment of the Auditor-General and a Panel of Experts to evaluate compliance with its order and file regular reports with the Court. In Black Sash II, the Court appointed a referee to make factual findings as to whether the Minister of Social Development had acted in bad faith when withholding a full disclosure of her conduct to the Court. This culminated in a personal costs order being issued against the Minister in Black Sash III.

This article offers an enriched account of these innovative remedies by juxtaposing an analysis of the Black Sash trilogy with similar developments in our lower courts. I focus on two particular lines of remedial development, reflecting the two major innovations featured in the Black Sash trilogy. First, I consider the emerging use of court-appointed agents to ameliorate supervisory jurisdiction, comparing the appointment of the Auditor-General and Panel of Experts in Black Sash I with the independent auditor in Madzodzo,⁵ the claims administrator in Linkside⁶ and the

¹ Black Sash Trust v Minister of Social Development & Others [2017] ZACC 8, 2017 (3) SA 335 (CC) (‘Black Sash I’); Black Sash Trust v Minister of Social Development (Freedom Under Law NPC Intervening) [2017] ZACC 20, (2017) 9 BCLR 1089 (CC) (‘Black Sash II’); Black Sash Trust (Freedom Under Law Intervening) v Minister of Social Development & Others [2018] ZACC 36 (CC) (‘Black Sash III’). I refer collectively to these decisions as ‘Black Sash’.


⁴ Madzodzo & Others v Minister of Basic Education & Others [2014] ZAECMHC 5, 2014 (3) SA 441 (ECM) (‘Madzodzo’).

⁵ Linkside & Others v Minister for Basic Education & Others [2015] ZAECGHC 36 (‘Linkside’).
NON-COMPLIANCE AS A CATALYST FOR REMEDIAL INNOVATION

special master in *Mwelase*. Second, the personal costs order in *Black Sash III* is evaluated as a strategy to ‘pierce the political veil’ and compared to the attachment of state assets in *Linkside*.

Although this turn to the lower courts is somewhat counter-intuitive for a journal focused on our apex court, drawing connections between these remedial developments foregrounds the cross-fertilisation between the different tiers of courts. Successful remedial experimentation in our lower courts can filter up to the Constitutional Court, while remedial precedents set by the Constitutional Court can be taken up by our lower courts. Tracing the roots of remedial innovation leads to a deeper understanding of these developments and fosters critical debate about the future direction of our remedial jurisprudence.

The cases covered in my analysis concern different constitutional rights – the right to social security in *Black Sash*, the right to basic education in *Madzodzo* and *Linkside*, and the right to security of land tenure in *Mwelase* – but they share a common remedial predicament. In each case, the litigation progressed in multiple stages as the court was confronted by the government’s continuing failure to comply with its positive constitutional duties. This prompted remedial escalation as structural orders became increasingly detailed and prescriptive through each stage of the litigation, culminating in a resort to novel remedial mechanisms in an effort to compel compliance. By juxtaposing accounts of the serialised litigation in *Black Sash*, *Madzodzo*, *Linkside* and *Mwelase*, the article captures this dynamic in which persistent non-compliance served as a catalyst for remedial innovation.

The argument proceeds as follows. Part II sets out the problem that my analysis addresses, namely a remedial predicament of persistent non-compliance. I contextualise *Black Sash* within the spate of recent Constitutional Court cases concerned with accountability for government compliance with constitutional obligations, and accordingly identify a need for greater judicial creativity in the exercise of constitutional remedial power. The heart of the article then considers two recent lines of remedial development: part III identifies the emerging use of court-appointed agents as an amelioration of supervisory jurisdiction, and part IV evaluates personal costs and the attachment of assets as strategies for ‘piercing the political veil’. I close in part V by briefly drawing out the implications of this account for our remedial jurisprudence.

II THE REMEDIAL PREDICAMENT OF PERSISTENT NON-COMPLIANCE

*Black Sash* and the other main cases covered in this article – *Madzodzo*, *Linkside* and *Mwelase* – all see the court confronted with persistent non-compliance in an extended remedial process that requires the government to implement systemic relief. The cases engage different

---

7 *Mwelase & Others v Director-General for the Department of Rural Development and Land Reform & Others* [2016] ZALCC 23, 2017 (4) SA 422 (LCC) (*Mwelase LCC*); *Director-General for the Department of Rural Development and Land Reform & Another v Mwelase & Others; Mwelase & Others v Director-General for the Department of Rural Development and Land Reform & Another* [2018] ZASCA 105, 2019 (2) SA 81 (SCA) (*Mwelase SCA*). At the time of writing, the appeal against the decision of the Supreme Court of Appeal was set down by the Constitutional Court for hearing on 23 May 2019.

8 I adopt this metaphor of ‘piercing the political veil’ from S Woolman ‘A Politics of Accountability: How South Africa’s Judicial Recognition of the Binding Legal Effect of the Public Protector’s Recommendations Had a Catalysing Effect That Brought Down a President’ (2018) 8 Constitutional Court Review 155, 185 fn 120. See further discussion at part II below.

9 Section 27(1)(c) of the Constitution of the Republic of South Africa, 1996 (‘Constitution’).

10 Constitution s 29(1)(a).

11 Constitution s 25(6).
constitutional rights – social security, basic education and security of land tenure – but they all concern failures by the government to fulfil those rights as required by the Constitution. The breaches have systemic impact on the rights of particularly vulnerable and disadvantaged groups: the fifteen million people reliant on social assistance; the thousands of children without desks or chairs or even teachers at some of the worst-off schools in the Eastern Cape; and the ten thousand labour tenants who do not enjoy secure rights over land that they and their forebears have farmed for generations. Where the government is required to take positive action to implement structural reform, non-compliance not only undermines the integrity of court orders and erodes respect for the rule of law, but also poses a systemic threat to rights. As non-compliance persists in these cases, court orders become increasingly detailed and prescriptive through each stage of the litigation, culminating in a resort to innovative remedial mechanisms to ensure accountability for full compliance. In short, non-compliance serves as a catalyst for remedial innovation.

Although this remedial predicament is perhaps more familiar to our lower courts than our apex court, accountability for compliance has emerged as a strong theme in recent Constitutional Court decisions. In the last few years, non-compliance with court orders has strained institutional comity in the context of corruption and abuse of state power, institutional dysfunction and disregard for the rule of law. Reflecting on two such cases, SABC v DA and EFF I, Stu Woolman argues that our apex courts have ‘become increasingly frustrated by the failure of the coordinate branches of government or organs of state to abide by even the most de minimus understanding of the rule of law’. He identifies a similar sentiment underlying the personal costs order against Minister Dlamini in Black Sash I:

12 AllPay Merits (note 2 above) at para 1.
13 This was the subject of litigation in Madzodzo (note 5 above). See Ready to Learn? A Legal Resource for Realising the Right to Education (Legal Resources Centre 2013)(‘Ready to Learn’) 51.
14 This was the subject of litigation in Linkside (note 6 above). See Ready to Learn (note 13 above) at 65–66.
15 The number of outstanding claims is recorded as being 10 914 in Mwelase LCC. Mwelase LCC (note 7 above) at para 9.
16 Woolman (note 8 above).
17 The Pandora’s box of accountability cases triggered by the Public Protector’s ‘Secure in Comfort’ report on impropriety in security measures installed at former President Jacob Zuma’s private Nkandla residence encompass the following matters: Economic Freedom Fighters v Speaker of the National Assembly; Democratic Alliance v Speaker of the National Assembly [2016] ZACC 11, 2016 (3) SA 580 (CC)(‘EFF I’); United Democratic Movement v Speaker of the National Assembly [2017] ZACC 21, 2017 (5) SA 300 (CC); Economic Freedom Fighters, United Democratic Movement, Congress of the People & Democratic Alliance v Speaker of the National Assembly & President Jacob Zuma (Corruption Watch as Amicus Curiae) [2017] ZACC 47, 2018 (2) SA 571 (CC)(‘EFF II’). See also South African Broadcasting Corporation v Democratic Alliance [2015] ZASCA 156, 2016 (2) SA 522 (SCA) (‘SABC v DA’).
18 Black Sash (note 1 above) is the obvious case in point. See also Pheko & Others v Ekurhuleni Metropolitan Municipality (No 2) [2015] ZACC 10, 2015 (5) SA 600 (CC)(‘Pheko II’).
19 Ignoring a High Court order for the arrest of visiting Sudanese President Omar Hassan al-Bashir is the most striking recent example of the government’s disregard for the rule of law: The Minister of Justice and Constitutional Development v The Southern African Litigation Centre [2016] ZASCA 17, 2016 (3) SA 317 (SCA); Southern Africa Litigation Centre v Minister of Justice and Constitutional Development & Others [2015] ZAGPPHC 402, 2015 (5) SA 1 (GP). South Africa’s notification to withdraw from the International Criminal Court halted the appeal to the Constitutional Court in the Al-Bashir case.
20 SABC v DA (note 17 above).
21 EFF I (note 17 above).
22 Woolman (note 8 above) at 175–176.
Holding individual ministers accountable – piercing ‘the political veil’ – reflects remarkable institutional confidence and indicates that the Court has grown weary of the contempt that government officials have shown for the law and their constitutional obligations. For the Court, these actions go to the very heart of the legal system’s legitimacy.

Confronted by intransigent non-compliance and a lack of accountability in political institutions, the Court has shown a greater readiness to grant intrusive remedies when it perceives the rule of law and the integrity of court orders to be at risk. In both Black Sash I and Mhlope, for example, the Court described the need for ‘extraordinary’ and ‘exceptional’ remedies to avoid a ‘constitutional crisis’.

While non-compliance has by no means ever been a rarity, the current trend stands in stark contrast to the good track record of compliance that had informed the Court’s (misplaced) optimism in Treatment Action Campaign that the ‘government has always respected and executed the orders of this Court. There is no reason to believe that it will not do so in the present case.’ The Court has often referred to its ‘broad remedial discretion’ to provide just and equitable relief, but its early reluctance to issue supervisory orders unless there were strong reasons to suspect non-compliance meant the breadth of these powers remained unexplored. Even when the Court began to rely on supervisory orders more frequently, they did not meet the kind of persistent and intransigent non-compliance that caused the remedial predicament in Black Sash.

Drawing on the principles laid down in Treatment Action Campaign, Roach and Budlender have developed a three-level model of remedies to serve as a guide for justifying remedial

---

23 Ibid at 185, fn 120.
24 The strain on institutional comity caused by non-compliance demonstrates Michael Bishop’s prescient insight that, ‘[u]nlike the more structural elements of the separation of powers, this respect can be lost and earned. If the executive continuously fails to comply with court orders, the Court will feel more comfortable issuing detailed interdicts or supervisory orders’. See M Bishop ‘Remedies’ in S Woolman & M Bishop (eds) Constitutional Law of South Africa (2nd Ed, 2006) Chapter 9 at 75.
26 The intervention in Black Sash I is described by the Court as ‘a remedy that must be used with caution and only in exceptional circumstances’ (Black Sash I (note 1 above) at para 51), while Mogoeng CJ similarly characterised Mhlope as ‘an exceptional case that cries out for an exceptional solution or remedy to avoid a constitutional crisis’ (Mhlope (note 25 above) at para 137).
29 Janse van Rensburg NO & Another v The Minister of Trade and Industry NO & Another [2000] ZACC 18, 2001 (1) SA 29 (CC) at para 28. See also Fose v Minister of Safety and Security [1997] ZACC 6, 1997 (3) SA 786 (CC) (‘Fose’).
30 With respect to the Court’s preference for a declaratory order over a supervisory order, see Rail Commuters Action Group & Others v Transnet t/a Metrorail & Others [2004] ZACC 20, 2005 (2) SA 359 (CC).
31 August & Another v Electoral Commission & Others [1999] ZACC 3, 1999 (3) SA 1 (CC); Sibiya & Others v Director of Public Prosecutions: Johannesburg High Court & Others [2005] ZACC 6, 2005 (5) SA 315 (CC); Nyathi (note 27 above). See also Bishop (note 24 above) at 189.
escalation. Their model charts the three reasons they identify as commonly underlying a government’s non-compliance with constitutional standards: inattentiveness, incompetence and intransigence. While their fundamental insight that each species of governmental failure calls for a different responsive technique is sound, the recent trend of non-compliance highlighted by this article suggests that their typology may need to be refined. Instead of framing non-compliance in psychological terms – inattentiveness, incompetence and intransigence – the cases featured in this article expose the institutional dynamic that often underlies non-compliance with constitutional obligations. In these cases, the retention of supervisory jurisdiction and the reliance on court-appointed agents are best understood as remedial mechanisms aimed at addressing institutional dysfunction and political blockages that threaten rights at a systemic level, rather than punitive measures targeting the recalcitrance of individual public officials.

This approach recognises that non-compliance with remedial orders does not only strain institutional comity but also undermines the normative commitment embodied by the right at stake, harming both successful claimants and similarly situated rights-holders. Remedies ‘concretise’ rights in the sense that they require courts to respond to social context in a way that gives practical and tangible effect to rights. It is in their crafting of remedies that courts most closely engage with the contextual realities of the state’s failure to fulfil its constitutional obligations, including the institutional problems that may underlie such a failure and jeopardise future compliance efforts. Furthermore, on a practical and immediate level, non-compliance severely depreciates the ‘cash value’ of a right for successful claimants, for whom the practical value of a right amounts to remedial relief that has tangible effect. Non-compliance therefore not only undermines the authority of court orders and the rule of law, but also harms the interests of rights-holders.

Section 172 of the Constitution empowers courts deciding constitutional matters to ‘make any order that is just and equitable’. Courts are therefore equipped with broad and discretionary remedial power to craft remedies that enhance accountability for compliance with constitutional obligations. Exercising this broad remedial power, courts need to look beyond the traditional remedial repertoire to find creative ways of overcoming the remedial challenges that threaten compliance with court orders. This imperative is captured by the oft-repeated call for remedial innovation issued by the Court in Fose to “forge new tools” and shape innovative remedies’ to ensure rights are vindicated.

---

33 Ibid.
34 Sandra Liebenberg has highlighted the importance of ‘responsive remedies’ that are attentive to contextual realities and promote transformative responses to socio-economic violations: S Liebenberg Socio-Economic Rights: Adjudication Under a Transformative Constitution (2010).
36 Pheko II (note 18 above) at para 27.
37 Constitution s 172(1)(b).
40 Fose (note 29 above) at para 69.
The wave of recent cases exposing a lack of accountability for government compliance with constitutional obligations has prompted the Court to shift gear in its exercise of constitutional remedial power.\(^{41}\) The Court has emphasised its strong constitutional mandate to enforce rights and the expansive framing of its constitutional remedial powers in discharging this mandate. In *Mhlope*, Mogoeng CJ elaborated on the nature of constitutional remedial power as follows in his judgment for the majority:

> Section 172(1)(b) clothes our courts with remedial powers so extensive that they ought to be able to craft an appropriate or just remedy even for exceptional, complex or apparently irresoluble situations. [...] If justice and equity would best be served or advanced by [a particular] remedy, then it ought to prevail as a constitutionally sanctioned order contemplated in section 172(1)(b).\(^{42}\)

Madlanga J’s separate judgment in *Mhlope* similarly registers this recent emphasis on the expansiveness of the Court’s remedial powers, warning that courts should not self-censor their exercise of these powers:

> The outer limits of a remedy are bounded only by considerations of justice and equity. That indeed is very wide. It may come in different shapes and forms dictated by the many and varied manifestations in respect of which the remedy may be called for. The odd instance may require a singularly creative remedy. In that case, the court should be wary not to self-censor. Instead, it should do justice and afford an equitable remedy to those before it as it is empowered to do.\(^{43}\)

These dicta signal the Court’s willingness to draw on its broad remedial powers to find creative ways of providing just and equitable relief. Moreover, given the recent line of cases in which the Court has ‘grown weary’\(^ {44}\) and ‘become increasingly frustrated’\(^ {45}\) by government non-compliance and disregard for the rule of law, it is likely the Court will more readily resort to innovative remedies that promote greater accountability for compliance with court orders in the future. Parts III and IV of this article identify two lines of remedial development which have emerged from litigation in our lower courts, and which may increasingly find favour with the Constitutional Court as they did in *Black Sash*.

### III AMELIORATING SUPERVISORY JURISDICTION: COURT-APPOINTED AGENTS

The first line of remedial development covered in this article concerns the emerging use of court-appointed agents as an extension or enhancement of a court’s supervisory jurisdiction. The retention of supervisory jurisdiction is now a well-established remedial mechanism used by our courts to supervise the implementation of a court order and thus ensure compliance with constitutional obligations.\(^{46}\) *Black Sash, Madzodzo, Linkside* and *Mwelase* exemplify the type of case which typically calls for supervisory jurisdiction, namely a breach of the state’s positive constitutional duties which requires the implementation of systemic relief. The forward-looking nature of positive duties poses a challenge to securing compliance because the implementation

---


\(^{42}\) *Mhlope* (note 25 above) at para 132.

\(^{43}\) Ibid at para 83.

\(^{44}\) Woolman (note 8 above) at 185, fn 120.

\(^{45}\) Ibid at 175.

\(^{46}\) *Pheko & Others v Ekurhuleni Metropolitan Municipality and Others (No 3) [2016] ZACC 20, 2016 (10) BCLR 1308 (CC)(‘*Pheko III*’) at para 1. See, generally, Bishop (note 24 above) at 196; Liebenberg (note 34 above); Viljoen & Makama (note 41 above); Roach & Budlender (note 32 above).
of systemic relief is subject to a range of temporal, contextual and institutional contingencies. Most obviously, compliance is contingent on positive steps being taken by those responsible for implementing systemic relief. This paradigmatically takes the form of a remedial plan to be implemented in stages rather than any once-off intervention. While conceptually distinct, a court’s task of crafting a remedy and its task of ensuring its enforcement are closely connected in such cases – there is a continuity between the remedy and its enforcement which may require sustained judicial involvement throughout an extended remedial process.

The retention of supervisory jurisdiction enables the parties to return to court without instituting fresh litigation in the event of non-compliance. This procedural function also has the “important symbolic effect” [of] interject[ing] a continued judicial presence in the affairs of the defendant that detracts from the defendant’s sense of institutional autonomy’.47 This background presence is the foundation of the court’s accountability-reinforcing role.48 However, supervisory jurisdiction also enables the court to engage in a more active oversight role during implementation to fulfil three important remedial functions. First, implementation may entail the prolonged and resource-intensive administration of systemic relief, such as the processing of large numbers of individual claims in a class action. Second, implementation needs to be subject to monitoring so that progress can be assessed and the remedial plan adjusted if need be. Third, reporting requirements can enhance transparency and accountability where there is a need for robust supervision over deadlines for implementing systemic relief.

Where these challenges of administration, monitoring and reporting strain the resources, capacity or competence of the court, comparative law demonstrates how the court may choose to rely on the parties, court-appointed agents or even civil society to assist with these supervisory functions. Public law litigation49 in the United States has led the way in the use of court-appointed agents to administer and supervise the implementation of systemic relief.50 Courts have drawn on a diversity of specialist functions in the form of special masters, monitors, mediators, administrators and receivers.51 Perhaps the most radical shift away from a court-centred approach to supervising the implementation of systemic relief is seen in India. The ongoing public interest litigation on the right to food is a noteworthy example of a truly collaborative approach to the implementation of systemic relief. Beginning with the landmark interlocutory decision of People’s Union for Civil Liberties v Union of India52, the Supreme Court’s active and sustained involvement in the implementation of its orders has been ‘both a response to and a catalyst for a well-organized, grass-roots activist campaign

50 For an excellent overview of the use of court-appointed agents in the United States, see Buckholz, Cooper, Gettner & Guggenheimer (note 47 above).
51 Ibid at 824–837.
52 Writ Petition (Civil) 196 of 2001.
of fact-finding, compliance monitoring, and strategic litigation’. While so-called ‘umbrella orders’ set out the lines of accountability for the implementation of its interim orders, the Supreme Court relies heavily on non-judicial mechanisms for promoting accountability for compliance. Court-appointed commissioners are assisted by a network of state-level advisors to monitor progress with the implementation of systemic relief and resolve disagreements with the government through deliberation.

The analysis below captures the nascent use of court-appointed agents in South African courts. In each case, the court-appointed agent fulfilled a specialised function that was tailored to the unique remedial challenges that emerged as the serialised litigation unfolded, but these functions share a common purpose in ameliorating the court’s supervisory jurisdiction. Where the court’s supervisory capacity was too limited to provide the necessary degree of oversight or expertise, the court-appointed agents served as independent monitors, auditors or administrators to ensure the implementation of systemic relief. I begin with an analysis of Black Sash I, before tracing its remedial roots to three cases brought by the Legal Resources Centre (LRC) in our lower courts: the use of an independent auditor in Madzodzo (cited as authority for the appointment of the Panel of Experts in Black Sash I), the claims administrator appointed in the Linkside class action, and the special master appointed by the Land Claims Court in Mwelase.

A A panel of experts: Black Sash

1 The remedial compromise in AllPay

The Black Sash trilogy takes place against the backdrop of the AllPay decisions, which concerned an unlawful contract concluded between SASSA and Cash Paymaster Services (Pty) Ltd (Cash Paymaster) to provide services for the payment of social grants for a period of five years. In AllPay Merits, the Constitutional Court found that the award of this tender to Cash Paymaster was constitutionally invalid, but grappled with the remedial consequences of this finding in a separate judgment, AllPay Remedy, after a further hearing on the question of remedy. The Court had to decide whether the unlawful tender should be set aside or whether ‘just and equitable relief’ called for a deviation from the ‘default position that requires the consequences of invalidity to be corrected or reversed where they can no longer be prevented’. In determining whether to depart from this default ‘corrective principle’, the Court had to navigate a tension posed by competing remedial imperatives: on the one hand, the need to ensure the rights of grant beneficiaries are protected through timely and uninterrupted payment of their social grants and, on the other hand, the need to provide administrative justice and uphold the rule of law in public procurement. In AllPay Remedy, the Court struck a compromise between these competing remedial imperatives, with Froneman J observing that

55 AllPay Merits (note 2 above).
56 AllPay Remedy (note 2 above).
57 Ibid at para 30.
58 Ibid at para 32.
59 Finn (note 4 above) at 261.
'a just and equitable remedy will not always lie in a simple choice between ordering correction and maintaining the existing position. It may lie somewhere in between'.

First, the Court sought to ensure the timely and uninterrupted payment of social grants by suspending its declaration of invalidity in respect of the contract concluded between SASSA and Cash Paymaster. It held that this exercise of discretionary remedial power falls squarely within the Court’s authority under s 172(1) of the Constitution, with its explicit textual support for ‘an order suspending the declaration of invalidity for any period and on any conditions, to allow the competent authority to correct the defect.’ In this case, the suspension served to protect the rights of grant beneficiaries rather than its usual function of preserving comity as a dialogic device that shows deference towards the other branches of government. By suspending the declaration of invalidity, the Court exercised its discretionary remedial power to ensure the continued operation of Cash Paymaster’s contractual operations and thus also the uninterrupted payment of social grants. Moreover, the Court held that Cash Paymaster assumed constitutional obligations when it entered into the social grants contract with SASSA, acquiring public power for the performance of a public function. This means that Cash Paymaster, functioning as an organ of state, ‘cannot simply walk away: it has the constitutional obligation to ensure that a workable payment system remains in place until a new one is operational’. Cash Paymaster’s continuing obligation to provide services for the payment of social grants therefore arose from both constitutional and contractual sources, with the latter being sustained through the Court’s exercise of remedial power in terms of s 172(1)(b)(ii) of the Constitution.

Second, the Court sought to vindicate the public interest and uphold the rule of law by ordering that the tender process be re-run. This order reflected the Court’s insistence that there be accountability and transparency in public procurement but did ‘not attempt to impose a final solution on SASSA’. Instead, recognising that SASSA was better placed to assess the impact of a new tender award on the range of interests implicated, the Court left it with the choice of either awarding a new five-year tender or taking over payment of social grants.
grants itself.\textsuperscript{67} Given the importance of the right to social security and the large number of beneficiaries affected by the public procurement of a social grant service provider, the Court maintained that the need for ‘disciplined accountability’ justified the imposition of a ‘structural interdict requiring SASSA to report back to the Court at each of the crucial stages of the new tender process’.\textsuperscript{68} The retention of supervisory jurisdiction to monitor SASSA’s progress was a forward-looking accountability mechanism aimed at securing prospective compliance once the suspension of the declaration of invalidity lapsed. In the brief judgment of \textit{AllPay III}, the Court clarified that it maintained supervisory jurisdiction to the exclusion of other courts until completion of the new tender process.\textsuperscript{69}

There can be no doubt that the \textit{AllPay} decisions represent a pioneering contribution to our jurisprudence. As du Plessis and Coutsoudis observe, \textit{AllPay} was the first occasion on which the Constitutional Court directly considered a review of a tender, setting out the appropriate review standard and emphasising the instrumental as well as intrinsic value of procedural fairness for guarding against corruption.\textsuperscript{70} Their praise rightly goes beyond the finding in \textit{AllPay Merits}, however. The bifurcated procedure in which remedial questions received independent and equal attention to the merits is welcomed as an ‘innovative’ and ‘refreshing’ approach worthy of replication in future cases.\textsuperscript{71} While remedy is too often consigned to a couple of sentences or paragraphs at the end of a judgment, frequently reading like a postscript or afterthought, \textit{AllPay} dedicates an entire judgment to remedial decision-making. The reasoning in \textit{AllPay Remedy} has been hailed as ‘laudable’,\textsuperscript{72} with its remedial compromise navigating ‘an inspired route via media between Scylla and Charybdis’\textsuperscript{73} that demonstrates ‘judicial creativity ... on full display’.\textsuperscript{74}

However, \textit{AllPay} comes with a sting in the tail. In November 2015, SASSA reported to the Court that it would not award a new tender but rather take over the payment of social grants itself from 1 April 2017. On the basis of this undertaking and the progress report filed in support of SASSA’s assurance that it would meet the 1 April 2017 deadline, the Court discharged its supervisory jurisdiction over the matter. On the eve of this deadline, however, it emerged that this assurance had not only been ‘without foundation’ but that the responsible functionaries of SASSA had in fact been aware for a year that it would not be able to comply with its undertaking to the Court.\textsuperscript{75}

\begin{enumerate}
\item \textbf{2. Justifying an escalated remedial response}
\end{enumerate}

SASSA’s non-compliance brought the country to the brink of a systemic breach of the social assistance rights of millions of people.\textsuperscript{76} In \textit{Black Sash I}, the Court was tasked with achieving ‘the practical avoidance of that potential catastrophe’ through the exercise of its discretionary remedial

\begin{itemize}
\item \textsuperscript{67} Ibid at paras 40–46.
\item \textsuperscript{68} Ibid at para 71.
\item \textsuperscript{69} \textit{AllPay III} (note 2 above) at para 16.
\item \textsuperscript{70} Du Plessis & Coutsoudis (note 4 above) at 764–765.
\item \textsuperscript{71} Ibid at 765–768.
\item \textsuperscript{72} Finn (note 4 above) at 258.
\item \textsuperscript{73} Du Plessis & Coutsoudis (note 4 above) at 768.
\item \textsuperscript{74} Ibid at 767.
\item \textsuperscript{75} \textit{Black Sash I} (note 1 above) at paras 5–6.
\item \textsuperscript{76} Ibid at para 43.
\end{itemize}
Given AllPay Remedy had sought to avoid continued reliance on the unlawful contract beyond the period of suspension, Froneman J expressed the Court’s frustration with SASSA’s intention to enter into a contract with Cash Paymaster without a competitive tender process:

This Court and the country as a whole are now confronted with a situation where the executive arm of government admits that it is not able to fulfil its constitutional and statutory obligations to provide for the social assistance of its people. And, in the deepest and most shaming of ironies, it now seeks to rely on a private corporate entity, with no discernible commitment to transformative empowerment in its own management structures, to get it out of this predicament.

Resigned to the inevitability of further tolerating an unconstitutional state of affairs, the question to be answered by the Court in Black Sash I was how the Court could enforce the performance of SASSA and Cash Paymaster’s self-acknowledged continuing constitutional obligations. As Froneman J observed, it was ‘for the Court in the exercise of crafting a just and equitable remedy to spell out the content of those obligations’.

Before delving into this remedial decision-making, however, the judgment began with a reflection on the nature and significance of the government’s non-compliance as the direct cause of the remedial predicament confronting the Court. SASSA and the Minister’s forced reply to the Chief Justice’s directions painted a picture of institutional dysfunction, with a diffusion of responsibility and shifting of blame among absent and transient incumbents of the office of CEO of SASSA. However, Froneman J’s analysis of the evidence goes beyond identifying the ‘demonstrated inability of SASSA to get its own affairs in order’ as the underlying reason for non-compliance. His characterisation uses language that suggests this incompetence was in fact coupled with intransigence. SASSA is said to have ‘walked away from the two fundamental pillars of this Court’s remedial order’ and ‘broken the promise in its assurance to the Court which formed the basis of the withdrawal of the supervisory order’. SASSA and the Minister of Social Development showed ‘no reciprocal comity’ towards the judicial branch and did not even ‘deign to inform the Court’ of its inability to meet the April 2017 deadline. The judgment of Black Sash I is described as the ‘judicial part of that [public] accounting in calling on government to explain its ‘conduct [that] puts grant recipients at grave risk and appears to disregard court orders’.

This searing indictment of SASSA’s conduct provides an important framing for the Court’s justification of its escalated remedial response. Most obviously, exposing SASSA’s intransigence allowed the Court to defend its order in AllPay Remedy and subsequent discharge of supervisory jurisdiction. Froneman J declared in absolute terms that the ‘conduct of the Minister and SASSA had created a situation that no one could have contemplated: the very negation of the

77 Ibid at para 15.
78 Ibid at para 8.
79 Ibid at para 41.
80 Ibid at para 48.
81 Ibid at para 9.
82 Ibid at para 19. See especially the answers in response to the Chief Justice’s first direction which attempts to clarify the person responsible on behalf of SASSA for ensuring compliance with AllPay Remedy.
83 Ibid at para 57.
84 Ibid at para 11.
85 Ibid at para 13.
86 Ibid at para 12.
87 Ibid at para 15.
88 Ibid at para 58.
purpose of this Court’s earlier remedial and supervisory order. His insistence that the Court had no reason to suspect non-compliance with AllPay Remedy was emphatic:

[T]here was no constitutional tension about social grants in November 2015. There was no legitimate reason for the Court not to accept the assurance of an organ of state, SASSA, under the guidance of the responsible Minister, that it would be able to fulfil an executive and administrative function allotted to it in terms of the Constitution and applicable legislation. There was no threatened infringement to people’s social assistance rights and no suggestion that the foundation of the Court’s remedial order would be disregarded. Now there is.

While it was arguably appropriate for the Court to defer to SASSA in the first instance, this characterisation is overstated as the government’s conduct in AllPay Remedy did in fact give some cause for concern about compliance. The Court had chastised SASSA in AllPay Remedy for having ‘adopted an unhelpful and even obstructionist stance’ and, citing this bad character reference, remarked in Black Sash I that, ‘[r]egrettably, not much has changed, except that this time around the Minister may have contributed to the continued recalcitrance’.

However, the insistence that non-compliance was not foreseeable added rhetorical force to the Court’s defence of the order crafted in AllPay Remedy and, more importantly, provided a basis for an escalated remedial response in Black Sash I. By foregrounding SASSA’s intransigence, the Court is able to cast its remedial intervention as one of reluctance and last resort:

It is necessary to be frank about this exercise of our just and equitable remedial power. That power is not limitless and the order we make today pushes at its limits. It is a remedy that must be used with caution and only in exceptional circumstances. But these are exceptional circumstances.

Everyone stressed that what has happened has precipitated a national crisis. Great care was therefore taken to frame the Court’s innovative remedial order as the necessary response to exceptional circumstances. The government’s intransigent non-compliance forced the Court’s remedial hand.

3 Resorting to a panel of experts

The first part of the remedial order in Black Sash I resembled the order in AllPay Remedy enforcing the reciprocal constitutional obligations between SASSA and Cash Paymaster for the payment of social grants. Drawing on its remedial powers under s 172(1)(b)(ii) of the Constitution, the Court declared that SASSA and Cash Paymaster are both under continuing constitutional obligations to fulfil the right to social assistance. They were accordingly mandated to ensure the payment of social grants under the same terms and conditions of the previous contract, subject to further safeguards for protecting the personal data of grant beneficiaries and auditing Cash Paymaster’s finances. The Court recognised, however, that just and equitable relief required more than simply identifying and circumscribing reciprocal obligations between SASSA and Cash Paymaster.

89 Ibid at para 36.
90 Ibid at para 10 (my emphasis).
91 AllPay Remedy (note 2 above) at para 75, quoted in Black Sash I (note 1 above) at para 56.
92 Black Sash I (note 1 above) at para 57.
93 Ibid at para 51.
94 Ibid order 4 at para 76.
95 Ibid orders 6 and 10 at para 76.
The second half of the remedial order in *Black Sash I* therefore features a range of remedial mechanisms to enhance accountability for SASSA’s compliance with its constitutional obligations. First, the Court held that SASSA’s failure to inform the Court that it would not be in a position to assume the payment of social grants from 1 April 2017 meant that the retention of supervisory jurisdiction alone ‘has been proved not to be enough to coax SASSA into doing what it was constituted to do. More is required.’\(^96\) Court supervision was therefore complemented by extensive reporting requirements, with both the Minister and SASSA being mandated to file regular progress reports with the Court detailing compliance with a timetable for deliverables.\(^97\)

Secondly, the Court introduced a bold remedial innovation to facilitate independent monitoring of SASSA’s compliance. Not content to rely solely on SASSA’s own evaluation of its progress, the Court accepted the argument put forward by the Black Sash Trust about the importance of independent court-appointed monitors.\(^98\) With none of the parties objecting to this proposal,\(^99\) the remedial order made provision for the appointment of the Auditor-General and suitably qualified independent legal practitioners and technical experts to jointly evaluate and report to the Court on SASSA’s compliance.\(^100\) This innovative remedy demonstrates institutional responsiveness to the reasons underlying SASSA’s non-compliance: the independence of the panel of experts offers much-needed accountability in light of SASSA’s intransigence, while their technical and legal expertise compensates for the institutional incapacity which has so far prevented SASSA from assuming payment of social grants itself.

While the Court spent considerable time in *Black Sash I* justifying the need for more intensive oversight over SASSA, surprisingly little was said about the theoretical basis for appointing a Panel of Experts or the significance of this innovation for our remedial jurisprudence. Three cases were cited as authority for this remedial intervention, although these references were consigned to a footnote without further explanation:\(^101\) *Grootboom*,\(^102\) *South African Human Rights Commission*,\(^103\) and *Madzodzo*. While all three cases affirm the importance of reporting and monitoring for ensuring compliance with structural relief, the appointment of an independent auditor in *Madzodzo* is different to the reliance on the South African Human Rights Commission (SAHRC) in the other two cases. In *Madzodzo*, as set out below, the independent auditor was a court-appointed agent charged with specific tasks as part of the court’s supervisory order. In the other two cases, by contrast, the SAHRC was already involved in the litigation in its capacity as an independent Chapter 9 institution, with the court in each instance simply recognising its constitutional duties and powers to monitor the observance of human rights in the country.\(^104\) In *Grootboom*, the Court simply observed that the SAHRC, as amicus, ‘will monitor and, if necessary, report in terms of these powers on the efforts made by the state to comply with its section 26 obligations in accordance with

\(^{96}\) Ibid at para 62.
\(^{97}\) Ibid orders 7–9 at para 76.
\(^{98}\) Ibid at para 71.
\(^{99}\) Ibid.
\(^{100}\) Ibid orders 11–12 at para 76.
\(^{101}\) Ibid at para 71, fn 37.
\(^{102}\) *Government of the Republic of South Africa v Grootboom* [2000] ZACC 19, 2001 (1) SA 46 (CC)(‘Grootboom’).
\(^{104}\) Constitution ss 184(1) and (2).
this judgment’. 105 The Court issued a declaratory order rather than a structural interdict, with no mention of the SAHRC’s reporting role since it was not part of a court-supervised remedial process. 106 In *South African Human Rights Commission*, the SAHRC was an applicant seeking (and being granted) an order requiring the respondents to provide them with regular reports setting out, among other things, the steps taken to comply with the court order on an ongoing basis. 107 The SAHRC in these cases was therefore not acting as a court-appointed agent as much as simply fulfilling its independent role as a Chapter 9 institution. The remedial roots of *Black Sash I* are therefore more properly found in *Madzodzo*. The next section recovers these roots by drawing out the resonances of *Madzodzo* with the remedial predicament faced in *Black Sash I*.

B An independent auditor: *Madzodzo*

1 Resorting to an independent auditor

*Madzodzo* concerns litigation brought by the LRC against the Department of Education to address the severe furniture shortages in public schools across the Eastern Cape. 108 It began in 2012 with a settlement approved by Griffiths J that was structured to address both the urgent furniture needs of the three individual applicant schools and the systemic shortage of furniture in schools across the province.

Besides requiring the Department to provide adequate desks and chairs to the three applicant schools, the court-approved settlement also set out a detailed plan for providing systemic relief. 109 First, the Department was directed to complete a comprehensive audit to record the furniture shortages of all schools in the Eastern Cape. 110 Second, the court order noted that the Department would ‘endeavour to ensure’ that the furniture needs established through the audit would be delivered to schools throughout the province by the end of June 2013. 111 This structured relief – involving an auditing stage and a delivery stage – recognised that the Department’s progress in implementing systemic relief cannot be monitored unless the scale of the problem is first ascertained. The remedial plan was fortified with extensive reporting requirements involving both the LRC112 and the affected schools,113 so that the Department’s progress with its audit and actual delivery of furniture could be tracked. Deadlines were attached to these reporting requirements.

In spite of the Department’s agreement with this plan for systemic relief, compliance was partial and begrudging. Although the individualised relief was implemented by duly delivering desks and chairs to the three applicant schools,114 little had been done by the Department to implement the structural relief. Moreover, much of what had been done in this regard was

---

105 Grootboom (note 102 above) at para 97.
106 Ibid at para 99.
107 South African Human Rights Commission (note 103 above) at paras 42 and 52.
108 Having undertaken extensive site visits over a period of several years, the LRC realised the enormity of the problem, noting that ‘thousands of children still sit on the ground because their classrooms have no, or an insufficient number of, desks and chairs. They hunch over workbooks and crane their necks to see the blackboard. They often get sick from sitting for hours on cold, dirty floors.’ *Ready to Learn* (note 13 above) at 51.
109 Madzodzo Court Order (Griffiths J, 29 November 2012).
110 Ibid at para 3.
111 Ibid at para 5.
112 Ibid at para 3.1 and para 7.
113 Ibid at para 4.
114 Madzodzo Supplementary Founding Affidavit at para 11.
either ineffectual or inadequate. The LRC was able to establish the extent of non-compliance when the Department submitted an incomplete and inaccurate audit of the province’s furniture needs.\textsuperscript{115} The Department had not made its task any easier by failing to issue a circular to inform schools of the audit and invite them to submit their furniture needs, as required by the court order.\textsuperscript{116} It was therefore unsurprising that many schools had been excluded from the audit and that the furniture needs recorded for those schools that were included in the audit had not been verified.\textsuperscript{117} Confidence in the accuracy of the audit was further undermined by glaring data irregularities, with similarities across school districts raising strong suspicions of data falsification.\textsuperscript{118}

The Department had failed to comply with the court order in other key respects too. It failed to present the LRC with a comprehensive plan setting out when it would deliver the furniture recorded in the (inaccurate and incomplete) audit,\textsuperscript{119} and had also failed to inform those schools due to receive furniture what they would receive and when it would be delivered.\textsuperscript{120} Another obstacle to implementation was the inadequacy of the furniture budget, with less than ten per cent of the required R360 million having been allocated to addressing the furniture shortage in 2013/2014.\textsuperscript{121} Having failed at this first stage of the plan for implementing structural relief, it was unlikely that the delivery phase of the plan would have been complied with, and indeed, by August 2013 no furniture had been delivered to schools using the R30 million budget that was in place.\textsuperscript{122} In short, it was clear that the Department had no reliable mechanism to ascertain or fund the actual furniture needs of schools in the province. Although the reporting requirements did not prevent non-compliance, they nevertheless functioned as a valuable diagnostic tool for detecting the extent and reasons for such non-compliance. By exposing these problems through regular reporting deadlines, the LRC was able to pick up on non-compliance relatively quickly and had clear grounds for returning to court.

The LRC filed an urgent application in August 2013 with the bold request that the Department be directed to appoint and pay an independent auditor.\textsuperscript{123} This remedial strategy was not newly envisaged by the LRC, as it had sought the appointment of an independent auditor during the early negotiations leading up to the initial settlement.\textsuperscript{124} The Department had strongly resisted this move at the time, however, claiming that an independent auditor was unnecessarily intrusive as it had already undertaken an audit which simply needed updating.\textsuperscript{125} Relying on this expression of political will by the government, the LRC conceded that an independent auditor need not be appointed. The inaccurate and incomplete audit subsequently presented by the Department led the LRC to regret this concession, as noted in the Supplementary Founding Affidavit during the second round of litigation in \textit{Madzodzo}:

\textsuperscript{115} Ibid at para 24.
\textsuperscript{116} Ibid at para 23.
\textsuperscript{117} Ibid at paras 31–32.
\textsuperscript{118} Ibid at para 33.
\textsuperscript{119} Ibid at para 25.
\textsuperscript{120} Ibid at para 46.
\textsuperscript{121} Ibid at para 12.1.
\textsuperscript{122} Ibid at para 12.2.
\textsuperscript{123} \textit{Ready to Learn} (note 13 above) at 52.
\textsuperscript{124} \textit{Madzodzo} Supplementary Founding Affidavit at para 54.
\textsuperscript{125} Ibid at para 55.
‘We negotiated in good faith with the respondents and granted their request to leave the audit process in their hands. This has proven a grave error.’  

In returning to court, the LRC therefore sought to tailor structural relief more effectively to both underlying reasons for non-compliance, namely the Department’s lack of capacity and its increasingly apparent lack of political will. In approving the terms of the second settlement, the High Court effectively agreed with the LRC’s argument that:

[T]he granting of a structured order and the appointment of an independent auditor is within the South African court’s power in this matter. Moreover, it is believed that such an order is warranted and necessary in light of the [Department]’s repeated violations of the immediately realisable, constitutional right to a basic education.

The order subsequently granted by Makaula J in September 2013 identified the specific tasks that the independent auditor would have to undertake in order to remedy the Department’s failed audit. This involved receiving reports from schools that had been left out of the audit and visiting the schools to verify their furniture needs. The Department was directed to file the revised audit at court and present a comprehensive plan for procuring and delivering the furniture as recorded by the independent auditor to all schools in the province.

2 Robust supervisory jurisdiction over deadlines

Following the appointment of an independent auditor to ascertain and verify the furniture needs of schools across the Eastern Cape, the LRC continued to monitor the implementation of systemic relief in Madzodzo closely. In spite of the appointment of the independent auditor, the institutional inertia in the Department continued to impede any meaningful progress from being made in meeting the furniture needs ascertained through the audit. By February 2014, children were still sharing desks and sitting on makeshift chairs such as beer crates, empty paint cans and sacks. A third round of litigation therefore ensued and the intransigence displayed by the Department received a sharp judicial rebuke. The judgment of the High Court in Madzodzo represents an important milestone in the LRC’s furniture litigation as it constituted explicit judicial recognition, in contrast to the implicit recognition of a court-approved settlement, that inappropriate or insufficient school furniture constituted a breach of the state’s positive duties in relation to the immediately realisable right to education.

The argument put forward by the Department in defence of their non-compliance with the second settlement was two-fold, and resembled the justifications frequently offered in relation to the failure to fulfil socio-economic rights: firstly, that budgetary constraints and the availability of resources had constrained their ability to satisfy the basic requirements of the right to education immediately and, secondly, that is was unreasonable to impose a fixed deadline for when the furniture needs of all public schools had to be met. In rejecting the first argument, Goosen J relied on the Constitutional Court’s holding in Blue Moonlight:

126 Ibid.
127 Ready to Learn (note 13 above) at 52.
128 Madzodzo Court Order (Makaula J, 26 September 2013) at para 4.1.
129 Ibid at paras 5 and 9.
131 Madzodzo (note 5 above) at para 20.
132 Ibid at para 22.
The court’s determination of the reasonableness of measures within available resources cannot be restricted by budgetary and other decisions that may well have resulted from a mistaken understanding of constitutional or statutory obligations. In other words, it is not good enough for the City to state that it has not budgeted for something, if it should indeed have planned and budgeted for it in the fulfilment of its obligations.133

The government’s failure to ensure that an adequate budget was made available for providing desks and chairs to all schoolchildren in the province was rejected as a defence for not giving immediate effect to the right to education. In respect of the Department’s second argument, Goosen J was scathing of the government’s resistance to a fixed deadline for furniture delivery because ‘the effect of the open-ended approach’ that they propose ‘offers little or no prospect that the furniture crisis will be addressed in the foreseeable future.’134 The judgment strongly holds that the immediately realisable nature of the right to education requires a clear timetable for the provision of relief.

Although the Department relied on these defences of budgetary constraints and impossibly short deadlines to claim that their non-compliance was not wilful,135 Goosen J maintained that these excuses ‘must be viewed against the backdrop of what has transpired since this application was first brought by the applicants in October 2012.’136 Contextualising the ostensible reasons for non-compliance within this longer perspective, he described their reliance on the unreasonableness of the deadlines in the settlement as ‘extraordinary in light of the fact that the terms of the order made by Griffiths J were negotiated between the parties and were accepted by the department’s officials as reasonable at the time that Griffiths J granted the order’.137 Goosen J therefore implied that the Department’s post hoc excuses exposed their lack of any genuine intention to comply with the terms of the settlement. He argued in light of the Department’s non-compliance that ‘this court is called upon to exercise its supervisory jurisdiction to ensure that the executive authorities charged with responsibility for ensuring the right of access to basic education act reasonably to fulfil their constitutional obligations’.138

The remedial order granted by Goosen J accordingly carved out a robust supervisory role for the court over clear deadlines for the implementation of systemic relief. The Department was directed to submit a revised independent audit to the court and the LRC, and to subsequently deliver all the furniture to schools in the province as recorded in the audit within 90 days.139 By setting this deadline, Goosen J refused to allow the Department’s preference for an open-ended approach to undermine the need for a clear timeline for the provision of systemic relief. At the same time, however, Goosen J fine-tuned the remedial order by softening the effect of the 90-day deadline for furniture delivery by allowing for the Department to apply for an extension of time:

To the extent however, that the exigencies of executing so significant a project may give rise to legitimate delays and therefore a legitimate inability to meet that projected time period, it will be appropriate to order that the time period may be extended at the instance of the respondents,

134 Madzodzo (note 5 above) at para 33.
135 Ibid at para 31.
136 Ibid at para 23.
137 Ibid at para 27.
138 Ibid at para 36.
139 Ibid orders 3 and 4 at para 41.
subject to full disclosure as to the steps already taken to meet the deadline and the projected time period within which the needs will indeed be met.\(^\text{140}\) The resulting order strikes a fine compromise between the insistence on a clear timeline for the provision of systemic relief and the value of flexibility to recalibrate the remedial order in response to unforeseen but legitimate implementation difficulties. This opening for the Department to apply for an extension of the deadline for furniture delivery should not be seen as a weakness in Goosen J’s remedy. Rather, recognising the likelihood that compliance would be patchy irrespective of any deadline, it sought to ameliorate the court’s supervisory jurisdiction over the deadline by requiring the Department to present a progress report and disclose the reasons for its non-compliance when it applied to the court for an extension. This prerequisite of a full disclosure was aimed at identifying the political blockages that were thwarting compliance.

The value of this approach for increasing transparency and accountability was demonstrated when, unsurprisingly, the Department requested an extension in May 2014.\(^\text{141}\) While predictably relying on their staple excuse of budgetary constraints, the Department’s justification for an extension also highlighted other key reasons for their partial compliance.\(^\text{142}\) Firstly, they had faced a number of legal challenges to their procurement processes, including litigation by the LRC on behalf of the Centre for Child Law concerning unlawful tender processes which resulted in delays in furniture delivery and a considerable waste of money.\(^\text{143}\) A second, but related, reason for the delays in furniture delivery was that the repeated irregular procurement processes resulted in the National Treasury assuming control of procurement, leaving the provincial Department in the position of being unable to purchase furniture.\(^\text{144}\) By requiring full disclosure before granting an extension, the court was able to exercise its supervisory jurisdiction in response to these obstacles to compliance.

With these institutional problems identified, a remedial plan could be developed to overcome them. In January 2016, the High Court ordered by consent that the Department appoint a school furniture ‘task team’ to prepare a consolidated list of the province’s furniture needs, which would then be verified and subsequently reconciled against deliveries to schools.\(^\text{145}\) Through the patient, robust and responsive exercise of supervisory jurisdiction, enhanced by the appointment of an independent auditor and, subsequently, a furniture task team, the systemic furniture needs of public schools across the Eastern Cape have been met.

C  A claims administrator: Linkside

The use of a court-appointed agent in Madzodzo met the need for an independent audit of the province’s school furniture shortages. Ascertaining the scale of the systemic problem was an essential first step in the remedial plan over which the court exercised robust supervisory jurisdiction. The Linkside class action featured an equally ground-breaking use of a court-appointed agent, but this time the particular function served was administering the large number of individualised claims involved in implementing systemic relief. The two cases

\(^{140}\) Ibid at para 40.

\(^{141}\) Fighting to Learn (note 130 above) at 37–38.

\(^{142}\) Ibid at 38.

\(^{143}\) Ibid.

\(^{144}\) Ibid.

\(^{145}\) Madzodzo Court Order (Brooks AJ, 26 January 2016).
therefore demonstrate different functions that a court-appointed agent might be required to fulfil in ameliorating the court’s supervisory jurisdiction.

The *Linkside* class action was the culmination of another stream of education litigation run by the LRC. This education litigation concerned the allocation and payment of teachers in public schools across the Eastern Cape. The LRC records that, at the time of first launching litigation on this issue in 2012, there were over 4 000 vacant teaching posts but also over 7 000 excess teachers in the Eastern Cape. This simultaneous shortage and surplus of teachers in the province revealed that the Department had significant difficulty complying with the national teacher post-provisioning norms that determine how many teachers should be employed and where they should be placed. The excess number of teachers holding posts at certain schools had the obvious consequence that the Department was overspending on teacher salaries. At the same time, however, schools with severe teacher shortages resorted to appointing temporary teachers at their own expense or through emergency donations from parents.

Over the course of protracted litigation that began with the case of *Centre for Child Law* and culminated in the class action of *Linkside*, the LRC sought to secure relief for schools whose teachers had not been formally appointed by the Department and therefore had not been paid. My discussion here is confined to the *Linkside* class action as it refined and extended the earlier remedial experimentation in *Centre for Child Law*.

The *Linkside* case itself unfolded in two phases. The first phase brought individual relief to particular schools and teachers listed as beneficiaries of the relief granted by the court. The novel aspect of this individualised relief was the use of so-called ‘deeming’ provisions, whereby a teacher who had been duly identified by the relevant School Governing Body to fill a vacant post and had been performing the functions of that post, would be deemed to be appointed. The use of ‘deeming’ provisions is discussed more fully in part IV below, as it facilitated the subsequent attachment of the Minister’s assets when the Department failed to pay those teachers who were deemed to be appointed. For the moment, it suffices to point out that the ‘deeming’ provisions were effective in providing relief for those individual teachers who were known to the LRC and could thus be identified and listed as beneficiaries of the relief approved by the High Court in the first phase of *Linkside*.

The scope of this relief was too narrow, however, as many teachers at similarly situated public schools did not benefit from the remedy. The first phase of the *Linkside* litigation therefore also launched an opt-in class action as a way of scaling up relief to all affected public schools in the province. In arguing for the certification of a class action, the LRC clearly articulated the rationale for the class action as being to secure systemic relief for the state’s failure to appoint and pay teachers:

> It is submitted that, in the absence of a class action mechanism, the claims of schools other than the applicants are unlikely to be enforced. Individual schools are unlikely to secure legal representation to bring their own claims for appointment and payment of educators. Each individual’s claim is

---

146 Ready to Learn (note 13 above) at 65.

147 Ibid.

148 Centre for Child Law & Others v Minister of Basic Education & Others [2012] ZAECGHC 60, 2013 (3) SA 183 (ECG)(‘Centre for Child Law’).

149 Fighting to Learn (note 130 above) at 54. The LRC intervened in its own name as an amicus curiae in Mukkadam v Pioneer (Pty) Ltd and Others in order to advocate for opt-in class actions, doing so with their litigation strategy in *Linkside* in mind. [2013] ZACC 23, 2013 (5) SA 89 (CC).
too small to justify litigation to enforce it. A class action is not only the most appropriate means to determine the claims, it is the only way to do so.\textsuperscript{150} The \textit{Linkside} opt-in class action was certified in March 2014, becoming the first class action of its kind in South Africa.\textsuperscript{151} It offered an opening for all public schools in the Eastern Cape affected by the government’s failure to appoint and pay teachers to join the proceedings and thereby receive relief. While 32 schools had initially approached the High Court requesting vacant permanent posts to be filled and a reimbursement of the R25 million they had paid in teacher salaries, this number rose to an additional 90 schools once the opt-in class action was certified and advertised.\textsuperscript{152}

The second phase of \textit{Linkside} concerned relief for all those schools and teachers who had opted into the class action. An important part of the remedial order granted by Roberson J concerned the payment and administration of outstanding teacher salaries. It declared that each of the specific amounts paid by the class member schools to teachers occupying vacant posts constituted a debt against the state.\textsuperscript{153} Roberson J recognised that this declaration alone was unlikely to provide effective relief, however, given the government’s poor track record in following through with these reimbursements.\textsuperscript{154} As covered in part IV below, the Department failed to reimburse schools for the outstanding salaries owed to those individual teachers identified in the first phase, with the result that state assets were attached to satisfy these debts. To avoid resorting to this radical intervention in the second phase of \textit{Linkside}, the LRC requested the appointment of a claims administrator to process the large number of claims by schools that had opted-in to the class action. Roberson J granted this request and appointed a claims administrator to oversee implementation of the court order.\textsuperscript{155} The Department was accordingly directed to appoint a firm of registered chartered accountants ‘to distribute the amounts payable to individual schools as members of the class and advise the court of their identity.’\textsuperscript{156} The Department was required to pay R81 million to the claims administrators so that they could verify each class member’s claim and disburse payments according to their respective entitlements.\textsuperscript{157}

Roberson J justified the appointment of a claims administrator on the basis of the Supreme Court of Appeal’s reference to court-appointed supervisors in \textit{Meadow Glen} as a remedial innovation that courts should consider in future.\textsuperscript{158} She pointed to the government’s poor track record of compliance in this matter and the added complexity of the class action as evidence for the ‘strong likelihood that administrative problems will be encountered in the implementation of payment by the Department’.\textsuperscript{159} She therefore held that ‘an efficient and independently accountable method of payment is essential’ to providing just and equitable

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{150} \textit{Linkside} Founding Affidavit at para 149.
\item \textsuperscript{151} See further J Rooney ‘Class Actions and Public Interest Standing in South Africa: Practical and Participatory Perspectives’ (2017) 33 \textit{South African Journal on Human Rights} 406.
\item \textsuperscript{152} \textit{Fighting to Learn} (note 130 above) at 54.
\item \textsuperscript{153} \textit{Linkside} (note 6 above) order 1.1 at para 1.
\item \textsuperscript{154} Ibid at para 20.
\item \textsuperscript{155} Ibid.
\item \textsuperscript{156} Ibid order 1.3.1 at para 1.
\item \textsuperscript{157} Ibid orders 1.3.2 and 1.3.3 at para 1.
\item \textsuperscript{158} \textit{Meadow Glen Home Owners Association & Others v City of Tshwane Metropolitan Municipality & Another} [2014] ZASCA 209, 2015 (2) SA 413 (SCA)('\textit{Meadow Glen}') at para 35.
\item \textsuperscript{159} \textit{Linkside} (note 6 above) at para 20.
\end{itemize}
\end{footnotesize}
relief. In prioritising the effective administration of payments, Roberson J dismissed the government’s objection about the cost implications of appointing a firm of chartered accountants, recognising that ‘their charges will not be insubstantial but it is unlikely they will make a significant inroad into the budget of the Department in its overall extent’.

The scale and complexity of relief in the Linkside class action meant that the High Court lacked the capacity and expertise to administer the processing of claims. The appointment of a claims administrator therefore ameliorated the court’s supervisory jurisdiction by providing the independent, expert and hands-on oversight that was required for the processing of mass claims against the Department. This role assigned to the claims administrator in Linkside class action was the first of its kind in our jurisprudence, but the important case of Mwelase represents a valuable opportunity for this innovation to be affirmed by our apex court as a tool for ameliorating supervisory jurisdiction where systemic relief entails the processing of mass claims.

D A special master: Mwelase

1 The need for a special master

Mwelase concerns the ongoing failure by the Department of Rural Development and Land Reform to process claims submitted by labour tenants seeking to obtain ownership of land they have worked on for generations but over which they were prevented from enjoying security of tenure as a result of past racially discriminatory laws and practices. Parliament enacted the Land Reform (Labour Tenants) Act 3 of 1996 (LTA) to enable labour tenants to gain security of tenure as promised in s 25(6) of the Constitution, but the Department has failed to comply with its constitutional imperative to process LTA applications. On its own admission, the Department seriously neglected applications between 2006 and 2015, even to the extent that it ceased processing claims altogether. This failure has perpetuated the tenuous position of over ten thousand labour tenants who have waited in vain to claim the entitlement s 25(6) promises them. Indeed, some of these claimants died before their claims were even gazetted. The litigation in Mwelase has itself been protracted, poignantly illustrated by the death of the first applicant, Bhekindlela Mwelase, on 7 November 2018.

The case began in July 2013, when the LRC launched litigation on behalf of the applicants in the Land Claims Court challenging the Department’s enduring and systemic failure to process LTA applications. Besides seeking individual relief for the first to fourth applicants, the LRC sought a structural interdict directing the Department to provide statistics detailing the status of all labour tenant claims and report regularly on its progress to the Land Claims Court until all outstanding claims had been settled or referred to the court for resolution. The Department repeatedly failed to provide the relevant statistics, in spite of having agreed to an order that it would collate and present the statistics to the court by March 2015. In April 2015, the Department estimated it would need a further two years merely to capture the basic details of the thousands of outstanding applications. Even before the Constitutional Court in 2019,

---

160 Ibid.
161 Ibid at para 21.
162 The deceased applicants are represented in the Constitutional Court by their heirs, but their tenuous position is aggravated by the argument put forward by the third respondent (the Hiltonian Society) in the Land Claims Court that their land tenure claims are not transmissible. I refer throughout this analysis to the applicants before the Land Claims Court and Constitutional Court as ‘the applicants’ even though they were of course the respondents before the Supreme Court of Appeal.
the applicants noted that this collation process still appears not to have been finalised. Similar to the challenge in *Madzodzo*, then, the Department was not able to ascertain the nature and scale of the problem, thus precluding progress from being made in addressing it.

Yet, as in *Madzodzo*, the reporting requirements served an important function in diagnosing non-compliance and prompting the Department to acknowledge its constitutional obligations to implement the LTA. With the poor state of the Department’s records exposed, the prospect of compliance was bleak in spite of the Department’s renewed commitment to processing applications. The applicants therefore considered the appointment of a special master to be the most effective form of supervision over a problem of this nature and scale. In June 2015, however, the parties were able to reach agreement on a detailed plan of court supervision that would enable the Land Claims Court to monitor the Department’s compliance. On the basis of regular reports, the court would be in a position to evaluate the targets set by the Department, assess its progress in reaching them, and recalibrate the implementation plan in the light of unexpected challenges that may arise during the remedial process.

However, this plan for court supervision failed on several fronts. The Department was unable to meet the deadlines set for filing its progress reports, it failed to comply with the implementation plan agreed to, and never responded to the concerns raised by the applicants in relation to the reports. The applicants returned to court with a renewed request for the appointment of a special master but, once again, a settlement was reached in an attempt to forge a mutually acceptable plan for implementation. This settlement agreement, approved by the Land Claims Court in May 2016, sought to foster good faith negotiation between the parties with a view to concluding a Memorandum of Understanding. As an alternative to a special master, the parties would establish a National Forum of NGOs to work together with the Department to implement the LTA. Instead of promoting co-ordination and co-operation between the parties, however, this negotiation agreement led to a breakdown in relations. The applicants claimed that the Minister of Rural Development and Land Reform had refused to negotiate in good faith, wilfully misconstruing the negotiation agreement by unilaterally establishing a National Forum in spite of vigorous objection by the applicants.\(^{163}\)

After two years of failed court supervision and multiple settlement agreements that met with non-compliance, the applicants returned to the Land Claims Court to push ahead with their request for the appointment of a special master. Although this move was a last resort after persistent non-compliance, the rationale for a special master put forward by the applicants remained unchanged: it offered the most effective means of supervision to co-ordinate the efforts of all parties in achieving the common goal of implementing the LTA. The appointment of a special master was therefore not a punitive measure but rather a mechanism directed at securing effective relief for the applicants.

In a comprehensive and carefully reasoned judgment handed down by Ncube AJ in December 2016, the Land Claims Court held that it was not only desirable but also urgent to enlist the aid of a ‘Special Master of Labour Tenants’.\(^{164}\) In reaching this conclusion, Ncube AJ found that ordinary court supervision had been undermined by the Department’s failure to

---

163 The Land Claims Court found that the Minister was not in contempt of the negotiation agreement. This finding was upheld by the Supreme Court of Appeal and is also the subject of the appeal pending before the Constitutional Court.

164 *Mwelase LCC* (note 7 above) at paras 33 and 36.
accurately set implementation targets or provide proper plans for reaching those targets. In setting out four key functions of the special master, the court was clear in characterising the role as being an agent of the court. First, the special master is an independent person who is appointed by, and reports to, the court. Second, his or her duty is to assist the court according to the mandate set by the court. Third, the special master’s decision-making powers are limited and always subject to court oversight. Finally, the special master enhances the court’s supervisory jurisdiction by bringing additional resources and specialised skills to bear on the case, engaging more comprehensively and informally with the parties than a judge is able to.

Based on this general characterisation of the nature of role of the special master, Ncube AJ identified the specific benefits to be gained from a court-appointed agent in Mwelase. On the one hand, the institutional dysfunction which had confounded the Department’s progress was said to ‘cry out for the intervention of a dedicated and competent person such as a Special Master’. The special master could aid the Department in developing a comprehensive strategy for the efficient processing of claims to ensure that the Director-General plays a proactive role in managing the referral process so that claims are adjudicated expeditiously. On the other hand, Ncube AJ held that a special master would significantly ameliorate the disadvantages that result from the capacity constraints suffered by the Land Claims Court. With only one permanent judge and a few acting judges each term, the Land Claims Court is ill-equipped to provide intensive, sustained court supervision over the implementation of systemic relief. In Mwelase, the eight different court appearances before the Land Claims Court had involved six different judges. Acutely aware of the court’s capacity constraints and the protracted litigation history in this case, Ncube AJ held that the appointment of a special master would be to the benefit of all concerned.

2 Misguided objections to a special master

On appeal, the Supreme Court of Appeal set aside the appointment of a special master. Writing for the majority, Schippers JA (Leach, Seriti and Willis JJA concurring) held that although it was ‘unassailable’ that the Department had breached its obligations in terms of ss 10, 25(6), 33, 195 and 237 of the Constitution, the appointment of a special master was ‘a textbook case of judicial overreach’. The majority’s objections to a special master extended beyond its necessity on the particular facts of Mwelase to embrace a principled stance against such an appointment ever being an appropriate exercise of remedial power.

In dealing with these objections, I argue that both the case-specific and principled reasons against the appointment of a special master are misguided. My critique is three-pronged, arguing that the majority judgment (1) failed to fully appreciate the remedial predicament faced by the Land Claims Court; (2) misconceived the function envisaged for the special master, prompting misplaced concerns about the separation of powers; and (3) ignored the

---

165 Ibid at para 17.
166 Ibid at para 19.
167 Ibid at para 29.
168 Ibid at para 27.
169 Ibid at para 28.
170 Ibid.
171 Mwelase SCA (note 7 above) at para 33.
172 Mwelase SCA (note 7 above) at para 51, quoting Mogoeng CJ’s separate minority judgment in EFF II (note 17 above) at para 223.
clear constitutional authority and precedential basis for the Land Claims Court’s true remedial discretion to appoint a special master. This critique is enriched by the compelling reasoning offered by Mocumie JA in her dissent.

The first cause for criticism of the majority judgment lies in its failure to fully appreciate the dynamic of protracted litigation and persistent non-compliance that culminated in the appointment of a special master. Schippers JA pointed to the earlier amenability of the parties to appointing a senior manager within the Department as evidence that a special master was unnecessary; he criticised the Land Claims Court for not launching an inquiry into non-compliance or asserting a more proactive oversight role; and he dismissed the Land Claims Court’s capacity constraints as providing justification for reliance on a special master. This characterisation of the litigation history implies that the special master offered a convenient outsourcing of the Land Claims Court’s supervisory function, rather than being the last resort after multiple alternatives had met with non-compliance. While the appointment of a special master had been raised by the applicants at the outset of litigation, they only pushed ahead with this request after both ordinary court supervision and negotiation had failed. The majority’s dismissal of the Land Claims Court’s own articulation of its limited capacity to engage in the kind of active oversight required in this case overlooks the protracted litigation history in this case – eight court appearances before six different judges over the course of several years. The serialised structure of the litigation in *Mwelase* resonates strongly with the tortuous trajectory of litigation in *Black Sash*, *Madsodzo* and *Linkside*, where in each case the court recognised the need for an independent court-appointed agent.

In contrast to the majority, Mocumie JA’s dissent is responsive to the broader context in which the litigation in *Mwelase* unfolded. She assessed the appropriateness of the special master in relation to the scale and urgency of what the majority conceded to be ‘unassailable’ constitutional violations. In short, the remedy was evaluated in light of the nature of the established breach. The Department’s track record of non-compliance extended beyond the confines of this litigation, as the claimants had been waiting 22 years for their applications to be processed. Mocumie JA was therefore careful to situate this ‘explicit and continued violation of constitutional obligations’ against the backdrop of our long and painful history of land dispossession and its perpetuation through the slow pace of land reform that our courts have lamented on numerous occasions.

Second, the majority misconceived the role that a special master would fulfil, giving rise to misplaced concerns about its implications for the separation of powers. Schippers JA described the appointment as ‘direct[ing] a complete outsider – the special master – effectively to take

---

173 Ibid at para 43.
174 Ibid at para 45.
175 Ibid at para 46.
176 Ibid at para 33 (majority) and para 71 (dissent).
177 Ibid at para 87.
178 Ibid at para 75. See further *In re Amaqamu Community Claim (Land Access Movement South Africa and Others as Amicus Curiae)* 2017 (3) SA 409 (LCC); *Land Access Movement of South Africa and Others v Chairperson of the National Council of Provinces and Others* [2016] ZACC 22, 2016 (5) SA 635 (CC)(‘LAMOSA I’); *Speaker of the National Assembly and Another v Land Access Movement of South Africa and Others* [2019] ZACC 10(‘LAMOSA II’). The pain – shame – of our colonial and apartheid history of land dispossession is most eloquently placed in constitutional context in *Daniels v Scribante and Another* [2017] ZACC 13, 2017 (4) SA 341 (CC).
over the functions and responsibilities of the DG and officials of the Department in relation to labour tenant claims’.\(^{180}\) In short, he understood the special master to be usurping executive power, fulfilling a function that ‘cuts directly across the powers of the DG’.\(^{181}\) The majority feared that this role would embroil both the special master and the Land Claims Court in the budgetary and operational allocations of the Department.\(^{182}\) This concern was the subject of Willis JA’s concurring judgment, which equated the appointment of a special master with the court ‘do[ing] battle with a potentially hugely expensive sword rather than a shield’.\(^{183}\) His reliance on this metaphor is telling, as it belies a bias towards shielding or protecting rights by enforcing the state’s negative duties of restraint, rather than boldly discharging the court’s constitutional mandate to provide effective relief where the state has breached its positive duties to fulfil rights.

The majority’s concerns about budgetary and operational allocations were largely speculative, based on a misconception of the functions the special master would fulfil. Its characterisation ignores the Land Claims Court’s careful delineation of the role of the special master in exercising limited powers according to the mandate set by the court and subject to judicial oversight.\(^{184}\) Indeed, reliance on a special master should not amount to judicial abdication. And, properly understood, this is not what the Land Claims Court ordered. On the contrary, Ncube AJ was alive to the potential friction that a special master could cause between the judiciary and the executive, and therefore applied his mind in amending the comprehensive draft order provided by applicants to ensure there would be no such intrusion in the functional domain of the executive.\(^{185}\) While the cost of a special master would likely not be trivial, it offers a more efficient form of judicial oversight in this case than ordinary court supervision, which is also a drain on judicial resources even if these costs are largely hidden. Given the capacity constraints of the Land Claims Court and the history of failed court supervision, the appointment of a special master would enhance the court’s supervisory jurisdiction. While cost may influence an assessment of what constitutes just and equitable relief in a particular case, such cost implications should be evaluated in light of what is required for the state to fulfil its constitutional obligations. As Mocumie JA recognised, ‘the doctrine of separation of powers is an important one in our democracy, but it cannot be used to avoid the obligation of a court to provide appropriate relief that is just and equitable.’\(^{186}\) The promise of s 25(6) of the Constitution cannot be realised cost-free, and the fact that the state has so far neglected or mistaken its obligations in this regard is no excuse.\(^{187}\) By indulging these excuses, the majority allows the state to have its cake and eat it, without paying.

The majority’s misconception of the special master’s role as being actual implementation, rather than oversight over implementation,\(^{188}\) was aggravated by a failure to distinguish the

---

180 Mwelase SCA (note 7 above) at para 48.
181 Ibid.
182 Ibid at para 50.
183 Ibid at para 98.
184 Ibid at para 19.
185 Ibid at para 37.
186 Ibid at para 89, relying on Fourie (note 62 above).
187 Blue Moonlight (note 133 above) at para 74.
188 Schippers JA relied on this distinction in finding that the applicants’ reliance on Fose (note 29 above) and Meadow Glen (note 158 above) was misplaced. See Mwelase SCA (note 7 above) at para 47.
diverse forms that this ‘foreign institution’ takes in the United States. While the functional role of a special master is primarily to assist the court in formulating rather than implementing the remedy, the majority’s characterisation is more akin to the intrusive roles performed by an administrator or a receiver, whose respective tasks are to supplement and replace the management of the relevant government institution. Moreover, as Mocumie JA recognised in her dissent, embracing the benefits of a special master does not require the wholesale adoption of a foreign institution but rather the careful adaptation of its core purpose to our unique jurisdictional setting and constitutional context:

...The social circumstances, historical reality of labour tenants, scope of powers of the LCC, specificity of our judicial methods to interpret transforming legislation and our courts’ ever available oversight powers would shape the institution of a special master in a way that makes it compatible, specific and appropriate in this context.

While foreign institutions should not be uncritically transplanted into South African law, a receptiveness to alternative approaches can spark creative solutions to shared challenges. This gestures towards the third shortcoming in the majority judgment, namely its disregard for the clear constitutional authority and precedential basis for the Land Claims Court’s discretionary exercise of remedial power to appoint a special master. The majority held that the appointment of a special master could not be justified on the basis of either the Land Claims Court’s authority to conduct informal or inquisitorial proceedings or its power to appoint a referee. Yet it failed to engage with the more obvious authority for the appointment of a special master to be found in the broad remedial powers granted to courts under s 172(1)(b) of the Constitution. No attempt is made to explain why the appointment of a special master falls outside the expansive scope of constitutional remedial power as affirmed in recent Constitutional Court cases like Mhlope, or to grapple with the implications of its decision for the prospect of securing effective relief for the applicants. While Schippers JA advanced a litany of objections to the appointment of a special master, only two paragraphs of his lengthy judgment are spent justifying the remedial order that replaces the Land Claims Court’s order.

Without the constitutional lens of s 172(1)(b), the majority’s scrutiny of the Land Claims Court’s remedial decision-making was out of focus. The correct perspective, as Mocumie JA set out in her dissent, is ‘whether the appellants have made out a case that justifies interfering with the Land Claims Court’s true discretion to grant appropriate remedies and to regulate its own process’. Interfering with a specialist court’s exercise of remedial discretion is only justified when there has been a ‘demonstrable blunder’ as a matter of legal principle rather than

---

189 Buckholz, Cooper, Gettner & Guggenheimer (note 47 above) at 826–837.
190 Ibid at 828.
191 Ibid at 831–837.
192 Mwelase SCA (note 7 above) at para 79.
193 Section 32(3)(b) of the Restitution of Land Rights Act 22 of 1994. See Mwelase SCA (note 7 above) at para 44.
195 For a matter dominated by remedial questions, it is noteworthy that no reference to s 172 of the Constitution is to be found in the Supreme Court of Appeal’s judgment.
196 Mhlope (note 25 above) as discussed in part II of this article.
197 Compare the objections to a special master in Mwelase SCA (note 7 above) at paras 38–54 with the justification of the order at paras 67–68.
198 Ibid at para 80.
judicial preference. The majority’s criticism, while forceful, was not correctly framed as a justification for interfering in the Land Claims Court’s exercise of true discretion to appoint a special master.

The final aspect of this third critique is borne out by this article’s contribution to foregrounding the rich precedential basis for court-appointed agents like a special master. Schippers JA dismissed the case law authority relied on by the applicants for the appointment of a special master and distinguished the Panel of Experts in Black Sash as having been tasked with only the very limited role of evaluating SASSA’s compliance. This functional distinction between the Panel of Experts in monitoring compliance and the special master in overseeing implementation is not misplaced, but misses the unifying purpose of both kinds of court-appointed agent in enhancing the supervisory jurisdiction of the court. By juxtaposing detailed accounts of Black Sash, Madzodzo, Linkside and Mwelase, this article has showcased the different functional roles that can be performed by a court-appointed agent – as monitor, auditor, administrator, special master – while simultaneously drawing out the shared remedial predicament that necessitated the resort to such remedial innovations. Instead of disregarding the lessons to be learnt from lower court litigation, it is hoped that the Constitutional Court’s decision in Mwelase will demonstrate a deeper understanding of pioneering cases like Madzodzo and Linkside that have relied on court-appointed agents to perform very similar functions to the role of the special master envisaged by the Land Claims Court.

IV PIERCING THE POLITICAL VEIL: PERSONAL COSTS AND ATTACHMENT OF ASSETS

The second line of remedial development identified in this article concerns a response to non-compliance that involves ‘piercing the political veil’. This metaphor seeks to capture the fact that public officials and government departments enjoy a measure of immunity as litigants before the court. When acting in their representative capacity, public officials are not usually affected personally by court orders against them. Instead, state resources and state assets are at stake. There are, of course, sound reasons for officials being spared exposure to personal liability for the performance of their representative functions, but there is a concomitant risk that this political veil may protect officials who are guilty of misconduct, gross negligence, incompetence or corruption from the court’s remedial reach. Court orders may be rendered ineffective where this conduct is the reason for persistent non-compliance with constitutional obligations. In this part, I consider two different strategies for ‘piercing the political veil’: first, an order of personal costs against a public official (costs de bonis propriis), as used in Black Sash against the Minister of Social Development; and second, the mechanism used in Linkside to formulate relief as an ascertainable debt against the state that can be satisfied through the attachment of state assets that benefit public officials, such as motor vehicles. These strategies offer innovative alternatives where contempt of court may be too blunt a tool, or may not be available because non-compliance concerns an order ad pecuniam solvendam (for the


200 Meadow Glen (note 158 above) at para 35.
payment of a sum of money) or because no officials can be individually identified as having been specifically tasked with the responsibilities in question.201

A Personal costs: Black Sash

While the award of costs is subject to judicial discretion, it has long been regarded a general rule that public officials should not be held personally liable for costs where, though mistaken, they acted *bona fide* in litigation concerning the performance of their official duties.202 Prior to *Black Sash*, the Constitutional Court recognised that a personal costs order against a public official was within its powers but was reluctant to do so in several cases where the issue arose.203 *Black Sash* stands as both a pioneering and paradigmatic case of ‘piercing the political veil’ through a personal costs order against a top-ranking government official. Not only was it the first time the Constitutional Court has issued costs *de bonis propriis* against a cabinet minister, but it also exemplifies the kind of conduct that can justifiably elicit this ‘mark of displeasure’ from the Court.204 Since the litigation history of *AllPay* and *Black Sash I* is covered in part III above, we need only set out here the course followed by the Court to establish whether the Minister of Social Development should be held personally liable for costs.

The issue of costs was left open in *Black Sash I* as the Court found that there were ‘reasonable grounds for investigating whether this Court’s remedial order was disregarded and, if so, whether this was done wilfully’.205 As the office-holder bearing primary responsibility for SASSA’s compliance with its constitutional obligations, the Minister was identified by the Court as the person who should be called upon to account for SASSA’s breach of its undertaking to the Court.206 The Minister was therefore mandated to file an affidavit with the Court explaining why she should not be joined in her personal capacity and why she should not be personally liable for the costs of the application.207

In her affidavit filed with the Court, Minister Dlamini argued against being joined to proceedings and bearing costs in her personal capacity. Her account instead laid the blame for the social grants crisis on officials from SASSA and the Department of Social Development, prompting two of these officials, the CEO of SASSA and the former Director-General of the Department, to enter the fray before the Court in *Black Sash II*.208 Their accounts contended that Minister Dlamini established parallel decision-making structures and lines of communication that bypassed officials in both SASSA and the Department, in contravention


203 Swartbooi v Brink [2003] ZACC 25, 2006 (1) SA 203 (CC) (‘Swartbooi’); Njungi v MEC, Department of Welfare, Eastern Cape [2008] ZACC 4, 2008 (4) SA 237 (CC). See also SA Liquor Traders Association v Chairperson, Gauteng Liquor Board [2006] ZACC 7, 2009 (1) SA 565 (CC) (‘SA Liquor Traders’). The costs award against the MEC was on an attorney-and-client scale rather than *de bonis propriis*. The state attorney was mulcted with costs on both an attorney-and-client scale and *de bonis propriis*.

204 *Black Sash III* (note 1 above) at paras 14 and 16.

205 Ibid at para 72.

206 Ibid at paras 73–74.

207 Ibid order 13 at para 76.

208 *Black Sash II* (note 1 above) at paras 1–2.
of government protocol. Faced with this conflicting affidavit evidence in *Black Sash II*, the Court joined the Minister as a party to the proceedings in her personal capacity and set out to disentangle the facts in order to establish whether she should be liable for costs in her personal capacity. Froneman J once again delivered judgment on behalf of a unanimous Court in *Black Sash II*. First, he drew on a range of constitutional values and principles to breathe new life into the common law rules for granting personal costs against a public official acting in a representative capacity. He identified accountable, responsive, transparent and effective government as the constitutional context within which to apply the test of bad faith and gross negligence to determine the Minister’s personal liability for costs. Second, the Court proposed resolving the conflicting factual accounts through the referral process provided for in s 38 of the Superior Courts Act. The parties subsequently agreed on a referee and the Court appointed the retired Ngoepe JP with a mandate to investigate the factual allegations at issue and report back to the Court with his findings. Following a full investigation, including testimony by Minister Dlamini, Ngoepe JP duly delivered his Inquiry Report to the Court and the parties were invited to file submissions as to whether Minister Dlamini should be held personally liable for costs. This reliance on a court-appointed referee not only assisted the Court in making factual determinations, but also revealed the potential for this referral process to enhance accountability through the public testimony and cross-examination of senior public officials like the Minister.

In *Black Sash III*, the Court characterised the Inquiry Report as ‘diplomatic but nevertheless damning’ and confirmed its essential factual finding that the Minister did not make a full disclosure to the Court for fear of being joined in her personal capacity and being mulcted personally in costs. Finding the inference of bad faith ‘irresistible’ in view of these factual findings, Froneman J described the Minister’s conduct as being at best ‘reckless and grossly negligent’. The Minister did not directly dispute the Inquiry Report but instead raised the belated objection that a personal costs order would offend the separation of powers. This last-ditch attempt to evade responsibility was roundly rejected on the basis of *Black Sash II*, with Froneman J again pointing out that the Minister’s culpable conduct was inimical to the Constitutional values she undertook to uphold as a member of the executive. The Court’s strong and unanimous stance against wilful non-compliance is clear from its insistence that ‘consequences must follow’ the Minister’s conduct in order to prevent similar disregard for

---

209 Ibid at paras 17–19.
210 Ibid at paras 3–4.
211 Ibid at paras 5–7.
212 Ibid at paras 7–9.
213 Act 10 of 2013. Ibid at para 23. This referral process in terms of s 38 of the Superior Courts Act is an example of a court-appointed agent as discussed in part III of this article, demonstrating yet another function that such an independent, court-appointed agent might perform to enhance the competence and capacity of the court in exercising its supervisory jurisdiction over a matter. See also *Pheko III* (note 46 above) at para 33, fn 33 and 34.
214 *Black Sash III* (note 1 above) at paras 2–3.
215 Ibid at para 4.
216 Ibid at paras 6 and 12.
217 Ibid at para 12.
218 Ibid at para 10.
219 Ibid at paras 10 and 14.
court orders in the future.\textsuperscript{220} The Court therefore took the unprecedented step of holding the Minister personally responsible for 20 per cent of the costs of litigation, sending a clear message that her conduct is ‘deserving of censure by th[e] Court as a mark of [its] displeasure’.\textsuperscript{221} Rubbing salt in the wound, the Court directed the Inquiry Report to be forwarded to the Director of Public Prosecutions for consideration as to whether Minister Dlamini should be prosecuted for perjury in light of the strong suggestion that she lied under oath in affidavits filed with the Court and in oral testimony before the referee’s inquiry.\textsuperscript{222}

\textit{Black Sash} bears out important principles underlying an order for costs \textit{de bonis propriis}. First, the Court’s description of a personal costs order as a ‘mark of displeasure’ is a feature it shares with other types of punitive cost orders. At the same time, however, personal costs are distinct from punitive cost orders made against officials in their representative capacity, as the former protects the public purse while the latter burden is ultimately borne by the taxpayer. The \textit{Biowatch} principle\textsuperscript{223} has gone a long way to counteract the inequality of arms in litigation involving the state, who as a public-subsidised litigant can afford to drag out litigation.\textsuperscript{224} Yet where public officials are indifferent to adverse costs orders against them in their representative capacity – even on a punitive scale – more is required. The personal costs order against such a high-ranking executive official in \textit{Black Sash} demonstrates the Court’s willingness to pierce the political veil where other forms of punitive costs orders are inadequate.

Second, \textit{Black Sash} affirms the constitutional basis for ordering costs \textit{de bonis propriis}. In our constitutional context, improper conduct could be grounded in either the institutional competence which can be expected of public officials or in the constitutional obligations which bind such officials.\textsuperscript{225} As Froneman J observed, there will be an overlap in most cases – where a personal costs order is justified on the basis of a public official’s conduct during litigation which falls short of the expertise and dedication expected of them, it is likely that the Constitution will also be vindicated through such an order.\textsuperscript{226} It is not entirely clear from the Court’s reasoning in \textit{Black Sash}, however, whether a public official’s breach of constitutional duties which merely gives rise to litigation but is not sustained through the course of litigation can, without more, justify a personal costs order. The Minister’s conduct before and during the litigation in \textit{Black Sash} rendered such a distinction unnecessary for justifying the Court’s personal costs order against her. For this reason, \textit{Black Sash} should not be taken as authority for ‘piercing the political veil’ merely because a court finds a breach of constitutional duties to be serious.

Indeed, courts should be wary of wielding the threat of personal costs orders to goad public officials into better performance of their constitutional obligations in future. As the Constitutional Court held in \textit{Swartbooi}, seeking to influence public officials in this way is an improper motive for ordering personal costs and trenches on the separation of powers.\textsuperscript{227} Instead, personal costs orders are more properly a way to mark the court’s displeasure with the

\textsuperscript{220} Ibid at para 13.
\textsuperscript{221} Ibid at para 14.
\textsuperscript{222} Ibid at para 17.
\textsuperscript{223} Biowatch Trust v Registrar Genetic Resources \& Others [2009] ZACC 14, 2009 (6) SA 232 (CC) (‘\textit{Biowatch}’).
\textsuperscript{224} Klaasen (note 201 above) at 618.
\textsuperscript{225} \textit{Black Sash II} (note 1 above) at para 8.
\textsuperscript{226} Ibid at para 8.
\textsuperscript{227} \textit{Swartbooi} (note 203 above) at para 25. See also Krüger (note 202 above) at 84.
conduct of public officials in connection with the litigation.\textsuperscript{228} The primary basis for ordering personal costs against a public official should therefore be understood as a breach of their constitutional duty to assist the court during litigation. Section 165(4) of the Constitution requires organs of state to assist and protect the independence, impartiality, dignity, accessibility and effectiveness of courts.\textsuperscript{229} This includes placing all relevant and material evidence before the court when engaging in litigation.\textsuperscript{230} As Black Sash demonstrates, a failure to disclose such information to the court provides a basis for a personal costs order.

This conduct may involve serious and sustained breaches of constitutional obligations during the course of litigation, particularly where court orders are met with persistent non-compliance, but it is distinct from punitive costs for the breach which gave rise to the litigation in the first place.\textsuperscript{231} This distinction is important because the punitive element of a personal costs order lies in it being a post facto response to conduct connected with the litigation, rather than a remedial measure aimed at curing the breach.\textsuperscript{232} It may sometimes be effective in targeting a political blockage within a non-compliant government department that is caused by the intransigence of particular public officials,\textsuperscript{233} but it is more often only a deterrent in the sense that it seeks to preserve the integrity of court orders and foster respect for the rule of law.\textsuperscript{234}

B Attachment of assets: Linkside

Where the political veil is a barrier to securing compliance such that a punitive, post facto personal costs order is premature or inappropriate, more creative remedial mechanisms may be needed. Linkside demonstrates one such remedial possibility for piercing the political veil to coerce compliance, namely the attachment of state assets. Until recently, s 3 of the State Liability Act 20 of 1957 prohibited the execution or attachment of state assets in satisfaction of a judgment debt against the government. This was a relic of our pre-constitutional order of parliamentary sovereignty that protected state officials from being held accountable for their failure to comply with court orders.\textsuperscript{235} In Nyathi, this form of state immunity was found to be unconstitutional and invalid. Section 2 of the State Liability Amendment Act 14 of 2011 now provides a mechanism for execution of state property.

\textsuperscript{228} This position accords with Innes CJ’s original formulation of the test as being whether the public official’s ‘conduct in connection with the litigation in question [was] mala fide, negligent or unreasonable’: see Vermaak’s Executor v Vermaak’s Heirs 1909 TS 679, 691 (emphasis added). Froneman J echoes this reasoning in Black Sash II. Black Sash II (note 1 above) at para 9 (‘[T]ests of bad faith and gross negligence in connection with the litigation, applied on a case by case basis, remain well founded’ (emphasis added)).

\textsuperscript{229} Matatiele Municipality v President of the Republic of South Africa [2006] ZACC 2, 2006 (5) SA 47 (CC) at para 107; SA Liquor Traders (note 203 above) at paras 48–49.


\textsuperscript{231} This distinction is pertinent for clarifying the basis of the High Court’s personal costs order against the Public Protector in Public Protector v South African Reserve Bank [2018] ZAGPPHC 175. At the time of writing, judgment is reserved in this matter before the Constitutional Court.

\textsuperscript{232} Klaasen (note 201 above) at 627.

\textsuperscript{233} MEC: Welfare (KZN) v Machi [2006] ZASCA 78 at para 11.


\textsuperscript{235} Nyathi (note 27 above) at para 18.
Linkside is a leading example of how this mechanism can ensure compliance where a government department lacks the institutional capacity and political agency to take the positive action necessary for providing relief. As discussed in part III above, the second phase of the Linkside litigation provided systemic relief through an opt-in class action in which a claims administrator was appointed to process claims against the Department of Education. The focus here, however, is on the first phase of the litigation, which provided individualised relief for 32 applicant schools by attaching state assets in satisfaction of outstanding teacher salaries. In order to facilitate the attachment of state assets, however, the relief first had to be formulated as an ascertifiable debt, which in turn required the relevant teachers to be duly appointed by the Department. The failure of the Department to appoint teachers was therefore the first obstacle to overcome, before the second obstacle of enforcing outstanding teacher salaries could be addressed through attachment.

Faced with the Department’s persistent failure to appoint or pay teachers, the LRC premised its remedial strategy on a default of non-compliance. Instead of requiring positive action by the government to appoint and pay teachers, the relief requested anticipated non-compliance by being designed to ensure that a failure to act would trigger the same result. This was achieved through a so-called ‘deeming provision’, whereby a teacher who had been duly identified by the relevant School Governing Body to fill a vacant post, and had been performing the functions of that post, would be deemed to be appointed. This had the paradoxical (but effective) result that the Department’s positive duties were enforced through inaction rather than positive action.

In March 2014, the Department agreed to the terms of settlement proposed by the LRC and the agreement was made an order of the court by Alkema J. The court-approved remedy ordered, firstly, that the 32 applicant schools be reimbursed for the money they had spent paying temporary teachers owing to the Department’s failure to appoint or pay them. These debts were declared to be enforceable in terms of the State Liability Act, as amended after Nyathi. The order anticipated non-compliance by specifying that, should these reimbursements not be paid within 120 days after the applicant submitted the details of the temporary teachers, the applicant schools were authorised to enforce these debts through execution under s 3 of the State Liability Act.

Secondly, the court-approved remedy declared that the temporary teachers currently filling those positions were deemed to be appointed permanently, and the Department was required to provide them with letters of appointment and pay them accordingly. Any outstanding salaries owed to these teachers deemed permanently appointed would also constitute ascertifiable debts subject to enforcement through attachment. While the subsequent opt-in class action scaled up relief to similarly situated schools, this first phase of Linkside confined relief to the lists of named teachers drawn up by the LRC in consultation with the 32 applicant schools. Only by individualising relief in this way could the outstanding teachers’ salaries constitute an ascertifiable debt enforceable through the attachment of state assets.

In September 2014, with the outstanding debts still not paid, the LRC issued the Department with a 14-day ultimatum after which time they would issue a writ to seize state assets to satisfy the R28 million debt owed to the 32 schools. Following through with this threat, the subsequent attachment of the movable assets of the Minister and top officials within the Education Department proved highly effective as the outstanding payments were made within days. Piercing the political veil in this way coerced compliance by targeting state assets that affected top public officials personally, such as the attachment of the Minister’s motor
vehicle. Unlike contempt of court or personal costs orders, the attachment of state assets does not require proof of *mala fides* on the part of individual public officials. As Ganesh observes in this regard, the attraction of attachment is that it is ‘relatively cold and clinical, and does not depend upon the composite mental state of the government department or its employees’. Whether non-compliance stems from intransigence or incompetence, attachment offers speedy and effective relief for judgment debts against the state.

V CONCLUSION

This article has showcased two lines of remedial development that are represented by the pioneering innovations in *Black Sash*: first, the nascent use of court-appointed agents to ameliorate supervisory jurisdiction and, second, the piercing of the political veil through personal costs orders against public officials and the attachment of state assets. In juxtaposing an account of *Black Sash* with similar remedial innovations featured in *Madzodzo*, *Linkside* and *Mwelase*, the article yields two particularly significant insights for the future of our remedial jurisprudence.

Firstly, the juxtaposition of *Black Sash* with lower court litigation foregrounds the cross-fertilisation in remedial development between different tiers of court. On the one hand, guidance for remedial innovation has filtered down from the Constitutional Court to our lower courts. The Court has issued clear authority for remedial development, identifying the ‘particular responsibility’ that obliges courts to “forge new tools” and shape innovative remedies where they are needed to provide effective relief. For example, both the Constitutional Court and the Supreme Court of Appeal have suggested the potential benefit of court-appointed agents, but much of the front-line experimentation in this regard is found in lower court litigation, as demonstrated by *Madzodzo*, *Linkside* and *Mwelase*. On the other hand, this remedial experimentation and innovation in our lower courts influences remedial development in the Constitutional Court. Some novel remedies get taken up by the Constitutional Court, often introduced by litigators or public interest organisations drawing on their experiences litigating in the lower courts. *Black Sash* and *Mwelase* offer leading examples of cases inspired by pioneering remedial experimentation in our lower courts. Much like comparative law analysis, an understanding of this cross-fertilisation can enrich our understanding of remedial developments and allow new ideas to take root.

Secondly, the juxtaposition of several streams of serialised litigation draws out their shared remedial predicament of persistent non-compliance which served as a catalyst for remedial innovation. These detailed accounts offer an insight into different manifestations of non-compliance in complex institutional settings of maladministration, dysfunction and even corruption. This suggests that the oft-cited typology of inattentiveness, incompetence and intransigence may need to be refined in order to fully capture the range and complexity of

236 Ganesh (note 201 above) at 33.
237 Ibid at 36.
238 Fose (note 29 above) at para 69.
239 In *Pheko III* (note 46 above) at para 33 footnotes 33 and 34, the Court pointed to the referral process provided for in the Superior Courts Act as a creative mechanism for enhancing supervisory jurisdiction, and alluded to the similarity between a referee and the American institution of a special master. Even more directly, the potential value of a special master was affirmed by the Constitutional Court in *Treatment Action Campaign* (note 28 above) at para 107 and by the Supreme Court of Appeal in *Meadow Glen* (note 158 above) at 35.
institutional reasons that might underlie non-compliance.\footnote{Inattentiveness, incompetence and intransigence characterise non-compliance in psychological terms that are most appropriately used in relation to individual states of mind or capacities. As this article demonstrates, however, non-compliance is more frequently the result of political blockages in public institutions that have chronically failed to meet their constitutional duties, with top officials often conceding as much.\footnote{A typology confined to inattentiveness, incompetence and intransigence can attribute responsibility to individual low-level officials based on discrete episodes of misconduct, but this often falls short of establishing systemic misconduct or the ultimate responsibility of senior public officials against whom relief is sought.\footnote{The remedial predicament identified in this article suggests that judicial responses to persistent non-compliance should primarily be targeted at the systemic functioning of institutions rather than the mental states of individual public officials.}

The innovations featured in this article gesture towards the future direction of our remedial jurisprudence. They underscore the breadth of constitutional remedial power that can be drawn on when, having ‘exhausted its lexicon of epithets’\footnote{In the case of MEC for the Department of Welfare v Kate [2006] ZASCA 49, 2006 (4) SA 478 (SCA) at para 29.} in attempts to spur the government into action, the court’s remedial hand is forced to forge novel mechanisms for compelling compliance. The value of remedial discretion is exemplified through such innovation, when courts do not resign themselves to the readily available repertoire of established remedies but rather imagine new possibilities for overcoming persistent non-compliance.\footnotetext[32]{See Roach & Budlender (note 32 above) at 345–351.\footnotetext[48]{Sabel & Simon (note 48 above) at 1062–1063.\footnotetext{Ibid at 1095.\footnotetext{MEC for the Department of Welfare v Kate [2006] ZASCA 49, 2006 (4) SA 478 (SCA) at para 29.}}}}
Invalid Court Orders

MITCHELL NOLD DE BEER

ABSTRACT: Over the past fifteen years, the Constitutional Court and Supreme Court of Appeal have held that objectively invalid administrative and executive decisions remain legally effective unless they are set aside in review proceedings. The rule has emerged from balancing two competing principles of the rule of law: on the one hand, legality in that government action must be lawful; on the other, certainty in that if government decisions could be ignored without recourse to the courts that could undermine the orderly functioning of government and the administration of justice. The courts have applied a similar rule in respect of court orders: all court orders are binding unless they are overturned on appeal or through rescission proceedings, save for one exception – where a judge issues an order outside of his or her authority or competence, it is invalid and not binding. In this article I trace the emergence and development of this exception in the jurisprudence and offer some preliminary justifications for why the courts treat invalid, authority-related court orders differently from invalid government decisions, thereby striking a different balance between legality and certainty. I then provide guidance as to how courts ought to determine whether a court order was issued with or without authority and explore the limits of the exception, particularly in respect of its application to orders issued by South Africa’s apex Court.

KEYWORDS: principle of legality, legal certainty, invalid decisions, judicial authority

AUTHOR: Pupil member of the Cape Bar. Email: mitchell.debeer@gmail.com

ACKNOWLEDGEMENTS: I am grateful for discussions with participants at the Constitutional Court Review IX Conference, where an earlier draft of the article was presented, and to the anonymous reviewers for particularly thorough and challenging comments. I am also grateful to Michael Bishop and Raisa Cachalia for their substantive and editorial input. Robert Freeman deserves special thanks for debating the views expressed in the article over the past few years, despite our disagreements. The usual warranties apply.
An important founding value of South Africa’s constitutional democracy is the rule of law. Derived from this founding value is the principle of legality, which requires that public functionaries may only exercise public power lawfully. Put another way, ‘people in positions of authority [must] exercise their power within a constraining framework of public norms.’ Unlawful government decisions are invalid and void, not voidable, and as an objective legal matter do not exist. Simple enough. But the rule of law requires much more. Certainty in the administration of justice is another principle of the rule of law. Citizens place reliance on extant government decisions and arrange their affairs around them, with the expectation that they were made lawfully and will be carried out. If government conduct, that might be unlawful, could simply be disregarded – without the need for a process to pronounce on its validity – that could lead to chaos and prejudice the rights and interests of many innocent people. The rule of law would suffer a blow if that were the case.

The tension between these two competing principles of legality and certainty has vexed our courts over the past two decades. Judges have had to figure out how to treat government decisions made in error by public functionaries but which have not been challenged in the usual way – judicial review proceedings. Should the court be entitled to ignore decisions made in error and not challenged properly? Or should such decisions be taken as valid or effective until so challenged?

The tension has bubbled to the surface most prominently in administrative law type cases. A long line of judgments – starting with the Supreme Court of Appeal’s decision in Oudekraal and most recently confirmed by the Constitutional Court in Magnificent Mile – have held that seemingly irregular administrative and executive decisions not challenged in review proceedings should be treated as legally effective unless and until set aside in the

---

1 Constitution of the Republic of South Africa, 1996 s 1(c) (“The Republic of South Africa is one, sovereign, democratic state founded on … [s]upremacy of the constitution and the rule of law.”)
The Court has emphasised that invalid administrative and executive acts while not existing in law, exist in fact and have consequences in the real world. If anyone could simply ignore a decision seemingly made in error (particularly its decision-maker) without the need for an independent judicial pronouncement on its validity that would be a licence to self-help. Legal certainty would be severely undermined. Cameron J explicitly refers to these as the ‘rule of law reasons’ which put a ‘provisional halt’ on the determination of a decision’s validity.

_**Oudekraal** and its application by the Constitutional Court have been controversial, and the rule it propounds has come under harsh criticism from a minority of the Court, which stresses the legality aspect of the rule of law. Nonetheless, a majority of the Court has consistently upheld and applied the rule, which has also been celebrated by many academics and court watchers. No doubt _Oudekraal_ provides for an important rule in our public law.

Finding the right balance between legality and certainty has also recently become controversial in respect of decisions made by other public functionaries – judges. Similar rule

---


7.  I specifically use the term ‘rule’ instead of ‘principle’ because _Oudekraal_ requires courts to treat unchallenged administrative and executive decisions conclusively as legally effective, which is subject to certain exceptions (eg, collateral challenges and perhaps instances of patent unlawfulness discussed in parts III and IV below). It is not a principle because its application does not lead to inconclusive outcomes in a given case. On the distinction between rules and principles, see R Dworkin _Taking Rights Seriously_ (1977) 17 and R Alexy _A Theory of Constitutional Rights_ (2002) Ch 3.

8.  _Merafong_ (note 6 above) at para 41 and fn 63.

9.  _Aquila Steel_ (note 6 above) at para 95.

10. Jafta J leads the charge with his allies Mogoeng CJ and Zondo DCJ together, on occasion, with the late Bosielo AJ. His reasoning, simply stated, is that Constitution s 172(1)(a) provides that a court ‘must’ declare conduct inconsistent with the Constitution to be invalid, ie he places the legality aspect of the rule of law above all else. Madlanga J formed part of the minority in _Kirland_ (note 6 above), but since then has supported the majority position in _Merafong_ (note 6 above), _Tasima_ (note 6 above) and even wrote the majority judgment in _Corruption Watch_ (note 6 above). For general discussions about this controversy, see J Brickhill, H Corder, D Davis & G Marcus ‘Administration of Justice’ in N Botha (ed) _Annual Survey of South African Law_ (2016) 3–12 and L Boonzaier ‘A Decision to Undo’ (2018) 135 _South African Law Journal _642, 659, 671–672.

of law issues arise in this context: ‘in exercising the judicial function judges are themselves constrained by the law’ and for good reason since judges make decisions which have far-reaching impacts on the rights and interests of ordinary people. A conundrum arises: while unlimited judicial authority is a constitutional oxymoron, it is for the judiciary to decide on the validity of all exercises of public power – including its own. The judiciary is both player and referee when it comes to this issue. The problem is that judges err sometimes. And if they err and their orders not challenged on appeal can simply be ignored, that could be a near-fatal blow to the rule of law; respecting court orders is a core foundation of our legal system. Not only will the law lose its ability to regulate affairs between private parties, but this might also undermine its important function in controlling the exercise of public power by the State.

Despite the majority of the Constitutional Court defending the Oudekraal rule at length in respect of administrative and executive decisions, the Court in Tasima and other cases seems to have accepted obiter that some court orders are not necessarily binding, even where they have not been challenged. Without needing to decide the issue, the Court has favourably referred to the Supreme Court of Appeal’s decision in Motala, which has developed a rule that if a court issues an order without ‘jurisdiction’ it is not binding, even if it is not challenged in the proper way – usually on appeal or in a rescission application. Call it the Motala exception.

In this article, I trace the emergence of the exception crafted in Motala and in so doing, try to make sense of and explain why our courts recognise this narrow exception by sometimes disregarding extant court orders that are made without authority, thus striking a different balance between legality and certainty. I then provide a blueprint for how exceptions of this kind should be determined and thereafter consider the limits of this exception, particularly in light of the difficulty of applying Motala to hard cases and to the decisions and orders of the Constitutional Court.

I AN OVERVIEW – THE MOTALA EXCEPTION

It is necessary upfront to provide a specific definition of a word used throughout the article and outline the basic structure of the exception. As for the definition, the decisions of the Supreme

13 L Kohn ‘The Test for “Exceptional Circumstances” Where an Order of Substitution is Sought: An Analysis of Trencon Against the Backdrop of the Separation of Powers’ (2017) 7 Constitutional Court Review 91, 93.
14 Jacobs & Others v S [2019] ZACC 4, 2019 (5) BCLR 526 (CC) at para 97 (Froneman J made this point quite candidly: ‘The italicised portion of that quotation [from an earlier judgment of the Constitutional Court] wrongly describes the law. Even Homer nodded. And courts sometimes make decisions per incuriam, or in a more brutal translation, “through lack of care”. The Latin phrase sounds more impressive than its English translation, but, embarrassing as it may turn out to be, one must examine whether the decision suffers from a lack of care.’)
Court of Appeal recognising the *Motala* exception refer to instances where a judge lacks ‘jurisdiction’ and that orders issued without ‘jurisdiction’ are considered to be invalid on their face. From a close reading of the jurisprudence, the term ‘jurisdiction’ is being used in a very specific sense – as a synonym for ‘power’, ‘competence’ or ‘authority’.\(^{17}\) To avoid confusion, since ‘jurisdiction’ may a have wider meaning in other contexts,\(^ {18}\) the word ‘authority’ is used in this article.

The basic rationale for the exception is stated as follows. A judge’s authority is found in the rules of law which stipulate the types of cases a court may hear and decide (‘subject-matter authority’) as well as the rules which specify the kinds of orders that the particular court is empowered to make (‘court-order authority’). The enquiry into whether a judge has ‘authority’ for these purposes, concerns not whether she exercises her adjudicative powers appropriately. Rather, it implicates the antecedent question whether the relevant powers are possessed in the first place – a narrower legal question. Whether a court has authority, so defined, to issue an order can be determined simply with reference to the order itself and the law (nothing further need be looked at in this regard). In this article, orders issued without authority are referred to as ‘invalid court orders’.\(^ {19}\)

To illustrate by way of example, whether a court exercised its authority appropriately by ordering the arrest of a debtor who was about to leave the country\(^ {20}\) is not the type of issue to which the exception is directed. Rather, the sort of issue is whether the court has the power to arrest such a debtor in the first place. This is a useful example because under the common law the courts had such authority, but it was declared unconstitutional.\(^ {21}\) Furthermore, the enquiry into a court’s ‘authority’ does not concern whether all the necessary facts were proved to sustain a cause of action, but whether, objectively speaking, the relevant court possesses the power to adjudicate on a claim of this kind. HLA Hart would have categorised these as a species of secondary rules – those which define and delimit the courts’ powers of adjudication as distinct from primary rules, which are intended to guide judges in the course of determining the lawfulness of the conduct in dispute.\(^ {22}\)

Based on the *Motala* exception, where a judge has decided a case or issued an order that falls outside of their adjudicative powers (ie, has made an authority-related error), the courts are entitled to treat it as invalid and not binding on its face – even if the order has not been taken on appeal.\(^ {23}\) Importantly, the courts have not, following *Motala*, extended this narrow exception to where a court makes other kinds of errors and issues an order. In fact, both the

---

\(^ {17}\) On how the word ‘jurisdiction’ is commonly used to refer to a decision-maker’s power or authority, see I. Baxter *Administrative Law* (1984) 302–303.

\(^ {18}\) H Wade & C Forsyth *Administrative Law* (11th Ed, 2014) 29–30 for a discussion about the wider use of the word.

\(^ {19}\) The Supreme Court of Appeal tends to call them ‘nullities’, but in *Tasima* the Constitutional Court rejected that terminology. See part II.C below.

\(^ {20}\) Referred to as an arrest *tanquam suspectus de fuga*.


\(^ {22}\) See generally, HLA Hart *The Concept of Law* (3rd Ed, 2012) ch V. Hart’s most famous secondary rule is his rule of recognition, by which decision-makers identify and ascertain primary rules. Since I borrow Hart’s characterisation, it is important to point out that he would not have supported the *Motala* exception (at 30).

\(^ {23}\) Of course one could also say that a judge acts outside of their authority or jurisdiction by issuing orders while making other kinds of errors. That is the essence of the wide *ultra vires* doctrine – see the discussion in Wade & Forsyth (note 19 above) at 29–30. For the purposes this article, however, ‘authority’ concerns only the narrow issue described above.
Constitutional Court and Supreme Court of Appeal have stressed the opposite: where a court has authority to decide an issue or make an order and errs, such an order must be treated as binding and effective unless overturned in the correct process. Admittedly, this is a subtle distinction that is not always easy to draw in practice.

A useful way of considering whether the Motala exception applies is to ask whether the court can ever issue the order, and not whether it may have been entitled to. The former is purely a legal question. If the answer is ‘Yes’ then the order is binding. If the answer is ‘No’ then the courts have held that it is not binding and can be disregarded.

II WHERE THE MOTALA EXCEPTION AROSE

The Supreme Court of Appeal, as mentioned in my introductory remarks, has had to confront the issue head on in cases before it, while the Constitutional Court has only considered it tangentially.

A The Supreme Court of Appeal’s approach

Motala concerned an order issued by Kruger AJ sitting in the High Court, purporting to appoint a judicial manager under the old Companies Act 61 of 1973. The Master of the High Court, who oversees these matters, refused to issue a certificate of appointment pursuant to the order, and appointed other insolvency practitioners as the provisional judicial managers. In subsequent proceedings in the matter, Legodi J of his own accord found the Master to be in contempt for refusing to follow the appointment order.

On appeal against the Legodi J order, Ponnan JA enquired into the validity of the earlier Kruger AJ order, even though it was not the subject of the appeal. Ponnan JA held that the latter order should not have been issued for want of authority. In coming to this conclusion, he surveyed jurisprudence developed by the Supreme Court of Appeal and its predecessor over the course of the previous century. He held that the High Court had no authority in law to make an order appointing a judicial manager in the first place. That power exclusively vested in the Master under South Africa’s statutorily-created insolvency processes, which Kruger AJ had impermissibly attempted to usurp.

I bracket that there is a second uncontroversial exception recognised in the jurisprudence to a court order being binding. Where an order has been issued against a party who has not been cited before the court, it will be a nullity (unless the order was taken in very special urgent circumstances). That was the position at common law (Lewis & Marks at 303). Constitution ss 34 and 35 entrench fair hearing and fair trial rights, which require that a party be given an opportunity to be heard before an adverse order is made against them (South African Riding for the Disabled Association v Regional Land Claims Commissioner & Others [2017] ZACC 1, 2017 (5) SA 1 (CC) at para 10).

24 I bracket that there is a second uncontroversial exception recognised in the jurisprudence to a court order being binding. Where an order has been issued against a party who has not been cited before the court, it will be a nullity (unless the order was taken in very special urgent circumstances). That was the position at common law (Lewis & Marks at 303). Constitution ss 34 and 35 entrench fair hearing and fair trial rights, which require that a party be given an opportunity to be heard before an adverse order is made against them (South African Riding for the Disabled Association v Regional Land Claims Commissioner & Others [2017] ZACC 1, 2017 (5) SA 1 (CC) at para 10).

25 Motala NO & Others v ITT Financial Corporation (Pty) & Others [2010] ZAGPPHC 162 at paras 14–15

26 Ibid at para 61.31.


29 Motala [note 16 above] at paras 5–6 citing old Companies Act s 429.

30 Ibid at para 14.
judge, Bertelsmann J in the same division had already handed down a judgment making this exact finding in law: only the Master has the power to make such appointments.31 Ponnan JA held that the Kruger AJ order was therefore a ‘nullity’. As a result, no finding of contempt could be made. Ponnan JA held so even though no proceedings to rescind or appeal against the Kruger AJ order had been brought: ‘Being a nullity a pronouncement to that effect was unnecessary. Nor did it first have to be set aside by a court of equal standing.’32

Another case is Changing Tides where a High Court judge hearing eviction proceedings issued an order compelling the Sheriff of the court to compile a list of occupants in the building concerned.33 Wallis JA34 declared those paragraphs of the order to be a ‘nullity’, as the Sheriffs Act did not confer such a power on the Sheriff – a creature of statute.35 He explained that ‘[t]hat part of the order was accordingly improvidently sought and erroneously granted’ and was ‘therefore a nullity.’36 The case has an important difference from Motala: the High Court did not take a power assigned to another actor by the law, but was effectively legislating from the bench by purporting to grant powers to the Sheriff that went beyond those provided for in the empowering statute.

In both matters the issue was not whether the court had authority to hear the dispute or whether the court was satisfied that facts were proved to justify the granting of the order. Rather it fell squarely within the realm of the antecedent question as to whether the court had the authority in law to grant the order in the first place.

B The Constitutional Court judgments

As for the Constitutional Court, a few recent cases touch on the issue. The first, Eke37 dealt with the following questions: whether the terms of a settlement agreement made an order of court are enforceable as a court order and what terms may legitimately be contained in such agreements.38 Madlanga J answered the first question in the affirmative: ‘Once a settlement agreement has been made an order of court, it is an order like any other’39 As to the second question, the Court held that not every settlement agreement, or every term of one, should automatically be made an order by a court, primarily for three reasons.40 First, the settlement agreement must relate to the issue between the parties that is before the court. Secondly, it must have some kind of ‘practical and legitimate advantage’ for the parties.41 And thirdly, most important for this discussion—

---

32 Motala (note 16 above) at para 14. Ponnan JA also proffered a sweeping claim that ‘it is a fundamental principle of our law that a thing done contrary to a direct prohibition of the law is void and of no force and effect’ (citing Schierbout v Minister of Justice 1926 AD 99, 109). This proposition is overbroad for focusing solely on the legality aspect of the rule of law. It overlooks that, for the sake of certainty, some unlawful acts might have to be taken as effective.
33 Changing Tides (note 16 above) at para 8.
35 Sheriffs Act 90 of 1986, s 3(1) sets out the functions of the office.
36 Changing Tides (note 16 above) para 9, citing Motala (note 16 above).
37 Eke v Parsons [2015] ZACC 30, 2016 (3) SA 37 (CC) (‘Eke’).
38 Ibid at para 1.
39 Ibid at paras 29 and 36.
40 Ibid at para 29.
41 Ibid at para 26.
the agreement must not be objectionable, that is, its terms must be capable, both from a legal and a practical point of view, of being included in a court order. That means, its terms must accord with both the Constitution and the law. Also, they must not be at odds with public policy.\textsuperscript{42} The Court did not explain whether the italicised part of the paragraph means that any error (whether authority-related or otherwise) by a court in deciding to make a settlement agreement an order will be binding.

A settlement agreement also played a central role in the second case, \textit{Tsoga Developers}. The applicants, officials from the North West Province, sought the stay of a writ of execution before the Constitutional Court pending the outcome of proceedings attacking the validity of the settlement. The agreement was alleged to have been made pursuant to an unlawful administrative decision, for want of compliance with s 217 of the Constitution.\textsuperscript{43} This notwithstanding, the High Court had refused to stay the writ. In considering whether the prospects of success in the pending review supported the granting of interim relief, the Constitutional Court discussed \textit{Motala} and \textit{Changing Tides}. Madlanga J wrote the judgment in this matter as well:

I read both judgments to say that, if on the face of the order, one is able to conclude that what the court has ordered cannot be done under the enabling legislation, the order is a nullity and can be disregarded. These cases are distinguishable from the instant scenario… On its face, [the settlement agreement] order is perfectly valid and competent. If there be a need to explain this, the so-called nullity of the settlement order does not – so to speak – \textit{jump out of the page}, as was the case with the nullity of the orders in \textit{Changing Tides} and \textit{Motala}. There has to be an antecedent step: proof of the grounds of review.\textsuperscript{44}

The applicants argued that because the settlement agreement had been made without complying with s 217, it did not ‘accord with both the Constitution and the law’ as required by \textit{Eke}.\textsuperscript{45} Madlanga J explained that \textit{Eke} was dealing with the requirements a court must be satisfied with \textit{before} making a settlement agreement an order of court, and that once it had done so ‘it is an order like any other’.\textsuperscript{46} In a footnote,\textsuperscript{47} he mentioned the Supreme Court of Appeal’s recognition of the possibility of an order being a ‘nullity’ in \textit{Motala} and \textit{Changing Tides}, but in any event made no ruling on the correctness of these judgments.\textsuperscript{48}

Note that in none of these cases did either of the courts extend the exception to invalidating orders arising out of other, non-authority related, errors. Interestingly, the old Roman Dutch authorities drew the same distinction. Voet for instance said:

By the customs of to-day such over stressful and pettifogging discussion of fine points of law as to whether a decision is \textit{ipso jure} void, or holds good by strict law and must be set aside through the remedy of an appeal, has been as far as possible abolished. The ruling has rather prevailed that decisions are never annulled under cover of nullity without appealing … \textit{There are exceptions when}

\textsuperscript{42} Ibid (emphasis added.)
\textsuperscript{43} Constitution s 217(1)(‘When an organ of state in the national, provincial or local sphere of government, or any other institution identified in national legislation, contracts for goods or services, it must do so in accordance with a system which is fair, equitable, transparent, competitive and cost-effective.’)
\textsuperscript{44} \textit{Tsoga Developers} (note 15 above) at para 50 (emphasis added.)
\textsuperscript{45} Ibid at para 52.
\textsuperscript{46} Ibid.
\textsuperscript{47} Ibid at fn 64.
\textsuperscript{48} Ibid.
the nullity arises from a lack of jurisdiction, or of summons, or of an attorney’s mandate as is noticed by the authors cited below following others.49

Again the reader will notice that the exception is narrowly confined to the three instances listed; the important one for my argument being ‘lack of jurisdiction’, which properly understood refers to a court’s authority. The first paragraph clearly states that all other orders are binding unless challenged properly.

C  Tasima

The next case decided by the Constitutional Court, Tasima complicates things and could be understood as putting the Motala exception in doubt. The case had convoluted facts and so it is necessary to set out its narrative at length properly to discuss the issues it dealt with.

The central issue in dispute was the legality and constitutionality of an agreement between the Department of Transport and a private company to create and operate a centralised, modern electronic road traffic management system.50 The company was paid generously by the State to provide these services. An acrimonious dispute regarding the validity of an extension of the agreement was referred to arbitration. Pending its outcome the company was granted an interim High Court order by Mabuse J in October 2012 (Interim Order), which kept the contractual obligations of each party alive and prohibited a breach of the agreement by the Department.51 The order was granted without enquiring into the validity of the extension.52

Despite the existence of the Interim Order and ongoing arbitration proceedings, the Department repeatedly attempted to take over the system. The company resisted and contended that this could only happen in terms of a convoluted takeover schedule to the agreement. Any premature succession, the company argued, would amount to a breach of the agreement and, by extension, the Interim Order as well. If the Department breached the agreement in any way, contempt of court proceedings were used as ‘the stick with which [the company] whipped the Department’s officials to submission’.53 A total of five contempt orders were granted between March 2013 and January 2014,54 entrenching the company’s position. Each order ruled that there had been contempt of the Interim Order and of the other contempt orders that had preceded it.

During the last iteration of contempt proceedings brought by the company in March 2015, seeking the committal to prison of the Department’s officials and those of the Road Traffic Management Corporation that had been joined along the way, the Department raised a new defence by impugning the validity of the agreement.55 According to the Department, the extension, on which the Interim Order was premised, came about in an unconstitutional manner for non-compliance with s 217 of the Constitution.56 The Interim Order was therefore invalid.

49 Voet 49.8.3 (trans Gane, vol 7) citing A. Matthaeus II De Auctionibus 1.16.59; Groenewegen De legibus abrogatis ad c 7.64; and Hollandsche Consultatien 5 c. 104 (emphasis added.)
50 Tasima (note 6 above) at para 6.
51 Ibid at para 32.
52 Ibid at para 33.
53 Ibid at para 43.
54 Ibid at paras 174–175. The orders were issued by Strijdom J, Ebersohn AJ, Fabricius J, Nkosi J & Rabie J respectively.
55 Ibid at para 48.
56 Ibid at para 49.
In raising the defence the Department did two things. First, it introduced a collateral challenge like a shield against the enforcement of the agreement through the order; and secondly, it launched a formal counter-application seeking that the extension be set aside, attacking it like a sword. Hughes J hearing this chapter of the story in the High Court upheld the Department’s counter-attack,57 thereby dismissing the contempt application.

The Supreme Court of Appeal was unconvinced.58 Brand JA held that the Department’s application to set aside the extension was unreasonably late by many years and, in exercising the court’s discretion, refused to entertain the review. The extension stood. It followed that the state actors were in contempt of the Interim Order, and of the contempt orders that followed it.59

On appeal, the Constitutional Court unanimously held it could decide the merits of the Department’s case despite the delay, concluding that the extension was unconstitutional and ought to be set aside.60 That left the question of contempt. On the one hand, Mabuse J had made the Interim Order to preserve the agreement, and judges following that had made findings of contempt for breaches of the Order. On the other, the very agreement underpinning the Interim Order had now been declared unconstitutional – and was therefore void from the outset. Should the Department be found to be in contempt of the orders? Jafta J for the minority held that on the facts of the case, the details of which are not pertinent for this article, the Department and its officials had not acted with contempt.61

For the majority, Khampepe J held that such an inference could not be drawn.62 In her view, the attempts by the Department and its officials to take over the system and transfer it to the Road Traffic Management Corporation was contemptuous of the Interim Order.63 She reasoned that up until Hughes J set aside the extension, the order was effectual, and that ‘the legal consequence that flows from non-compliance with a court order is contempt.’64 Khampepe J held that not complying with the Interim Order from the time it was made until the time the High Court set aside the extension of the agreement was contemptuous, but that committal was not appropriate on the facts.

In support of these conclusions Khampepe J relied on three bases. First, she drew on s 165(5) of the Constitution and reasoned that although the Constitution entrenches its

57 Tasima (Pty) Ltd v Department of Transport [2015] ZAGPPHC 421.
59 Ibid at para 20.
60 Although unanimous in this conclusion, the Court was bitterly divided in its reasoning, with a primary disagreement about the Oudekraal rule. Khampepe J for the majority (Froneman J, Madlanga J, Mhlantla J and Nkabinde J concurring) confirmed the rule again. She held that even though the counter-application was brought unreasonably late by the Department, it should nevertheless be entertained in the circumstances. The extension was found to be invalid and set aside. Jafta J for the minority (Mogoeng CJ, Bosielo AJ and Zondo J concurring) held that a court is always obliged to set aside unconstitutional government decisions, no matter if they were formally challenged on review or raised collaterally in enforcement proceedings. He also criticised the Oudekraal rule and the majority decision in Kirland. Froneman J in a concurring majority judgment criticised Jafta J for ignoring precedent in his reasoning. Zondo J in turn wrote a concurring minority judgment which criticised Froneman J for doing the same thing in previous cases. The Court also included a single paragraph judgment relating to a concern raised by one of the parties in the case, that Jafta J had prejudged the matter for having referred to the factual situation in Tasima as an example in his minority judgment in Merafong (note 6 above), which had been handed down a few weeks before.
61 Tasima (note 6 above) at paras 114–115.
62 Ibid at para 184.
63 Ibid at para 185.
64 Ibid at para 186.
supremacy in s 2,65 it expressly provides for one exception to a court order’s immediate binding force – confirmation of an order of constitutional invalidity by the Constitutional Court.66 This ‘tipped’ the balance in favour of holding that an ‘order is binding, irrespective of whether or not it is valid, until set aside.’67 Secondly, she buttressed this proposition by referring to pre-constitutional jurisprudence.68 Thirdly, Khampepe J applied the Oudekraal rule by analogy to invalid orders:

This reading of section 165(5) accepts the Judiciary’s fallibilities. As explained in the context of administrative decisions, “administrators may err, and even … err grossly.” Surely the authors of the Constitution viewed Judges as equally human. The creation of a judicial hierarchy that provides for appeals attests to this understanding. Like administrators, Judges are capable of serious error. Nevertheless, judicial orders wrongly issued are not nullities. They exist in fact and may have legal consequences.69 This conclusion ‘vindicates the constitutionally-prescribed authority of the courts’. Khampepe J cautioned further that ‘[a]llowing parties to ignore court orders would shake the foundations of the law’70 and continued that ‘[t]he duty to obey court orders is the stanchion around which a state founded on the supremacy of the Constitution and the rule of law is built.’71 Failure to comply with a court order thus results in contempt.72

It is in this context that the majority took issue with some of Ponnan JA’s reasoning in Motala.73 Khampepe J analysed and placed in doubt the Supreme Court of Appeal’s reliance on its century of jurisprudence74 and criticised it for failing to mention s 165(5) in its judgment.

One reading of the decision might be that the exception is put in doubt.75 The reasoning of the Court seems to imply that court orders should always, as a rule, be considered binding and capable of founding contempt, until set aside.76 Nonetheless, it is important to emphasise that the Court was not being called upon to consider the status of an order made without authority as defined in this article. The question in the matter was whether Mabuse J – who issued the Interim Order – ‘had the authority to make the decision that he did at the moment that he

---

65 Constitution s 2 (‘This Constitution is the supreme law of the Republic; law or conduct inconsistent with it is invalid, and the obligations imposed by it must be fulfilled.’)
66 Constitution s 172(2)(a) (‘The Supreme Court of Appeal, the High Court of South Africa or a court of similar status may make an order concerning the constitutional validity of an Act of Parliament, a provincial Act or any conduct of the President, but an order of constitutional invalidity has no force unless it is confirmed by the Constitutional Court.’)
67 Tasima (note 6 above) at para 180. (Khampepe J contends as follows at fn 118: ‘Even though courts do not have the purse or sword to enforce their orders, the effect of their decision is binding in law.’)
68 In re Honeyborne (1876) 7 Buch 145; S v Zungo 1966 (1) SA 268 (N).
69 Tasima (note 6 above) at para 182 citing Kirland (note 6 above) at para 90.
70 Tasima (note 6 above) at para 183.
71 Ibid.
72 Ibid at paras 184–187.
73 Ibid at para 190.
74 Ibid at paras 188–196.
75 For example, the editors of the South African Law Reports state that Motala was ‘criticised’ by the Constitutional Court in Tasima. See also D van Loggerenberg Erasmus Superior Court Practice (Revision Service 6, 2018) A2-94B–95 (Notes that Tasima takes a different approach from Motala.)
76 In an earlier draft of the article I adopted the view that the exception carefully drawn in Motala was seriously put in doubt by the reasoning in Tasima: any order binds, but judges have a discretion not to enforce one made without authority. I am indebted to both of the CCR’s anonymous reviewers as well as my editors for suggesting that I reconsider that view, as I have done.
made it.\textsuperscript{77} Undoubtedly, he had the power to issue an interim order to maintain the status quo pending the determination of the agreement’s validity in the arbitration proceedings.\textsuperscript{78} Khampepe J explicitly stated that \textit{Motala} was ‘dealing with a different issue’\textsuperscript{79} from \textit{Tasima}.\textsuperscript{80} Again, buried in a footnote, one finds the Court recognising the exception: ‘\textit{Motala} correctly holds that where an order is made without jurisdiction … another court may refuse to enforce it’.\textsuperscript{81} In doing so, the Constitutional Court adds a wrinkle. Khampepe J explicitly stated that invalid orders are not ‘nullities’, but emphasised that a court can still disregard an earlier order made without authority, albeit in a footnote.\textsuperscript{82}

Notably, the Supreme Court of Appeal in \textit{Moraitis Investments} subsequently interpreted \textit{Tasima} as having confirmed the \textit{Motala} exception. After stressing that \textit{Tasima} underscored the general proposition that court orders shouldn’t be ignored, Wallis JA cited the footnote mentioned above for the proposition that ‘[t]here is a narrow exception where a court makes an order that is on its face beyond its powers’.\textsuperscript{83} This too was contained in a footnote.

\section{WHY TREAT COURT ORDERS DIFFERENTLY FROM ADMINISTRATIVE AND EXECUTIVE DECISIONS?}

I have already outlined some of the ‘rule of law reasons’ why South Africa’s apex courts have insisted on treating unlawful government decisions as effective until set aside. Surely some of the same reasons apply to court orders as well. Regarding those reasons, the courts have held that just because something is invalid in law, it should not necessarily be treated as a nullity in reality because it exists in fact and may have legal consequences. This distinction between the normative legal validity of an act on the one hand, and the same act’s factual existence and effect on the other, was drawn by Christopher Forsyth.\textsuperscript{84} In \textit{Oudekraal}, the Supreme Court of Appeal approvingly cited Forsyth’s work in developing the rule.\textsuperscript{85} Additionally, the courts are anxious to diminish the risk of public functionaries resorting to self-help in deciding which of their decisions are lawful or not — something that is the province of the courts.\textsuperscript{86}

\begin{thebibliography}{99}
\item \textsuperscript{77} \textit{Tasima} (note 6 above) at para 198.
\item \textsuperscript{78} Arbitration Act 42 of 1965 s 21(f) empowers the court to make interim orders of this nature.
\item \textsuperscript{79} \textit{Tasima} (note 6 above) at para 197.
\item \textsuperscript{80} Perhaps the majority was responding to sentiments expressed by Jaftha J in his lone judgment of \textit{Nkata} (note 15 above) that other non-authority related errors may also make court orders nullities, which is not consonant with the \textit{Motala} exception.
\item \textsuperscript{81} \textit{Tasima} (note 6 above) at fn 156.
\item \textsuperscript{82} If we understand \textit{Tasima} in this way, what then do we make of Constitution s 165(5)? It provides that an ‘order or decision issued by a court binds’? Perhaps \textit{Motala} and \textit{Tasima} mean that a ‘court order’ issued without authority ought not even be referred to as a ‘court order’ at all, for it does not meet the desiderata for an order. On this approach, an ‘order or decision’ for the purposes of Constitution s 165(5) only refers to those made with authority.
\item \textsuperscript{83} \textit{Moraitis Investments} (note 16 above) at fn 4.
\item \textsuperscript{84} C Forsyth ‘“The Metaphysic of Nullity” Invalidity, Conceptual Reasoning and the Rule of Law’ in C Forsyth and I Hare (eds) \textit{The Golden Metwand and the Crooked Cord: Essays in Honour of Sir William Wade QC} (1998) 147. Forsyth draws on Hans Kelsen’s pure theory of law for the theoretical foundation of the distinction explaining that the theory ‘is built on the distinction between the \textit{Sein} (the Is) and the \textit{Sollen} (the Ought), between the realm of things that are – facts or natural phenomena – and the realms of norms, included therein the law.’
\item \textsuperscript{85} \textit{Oudekraal} (note 6 above) at para 29.
\item \textsuperscript{86} \textit{Kirland} (note 6 above) at para 103.
\end{thebibliography}
There are further considerations as well. It has always been the case that the judicial review of an unlawful decision must be brought within a reasonable time and by a party with sufficient interest in the decision to challenge it. The courts have always had a discretion in deciding whether or not to set aside an unlawful decision in review proceedings. For these reasons it has been said that ‘the court will invalidate an order only if the right remedy is sought by the right person in the right proceedings and circumstances.’ Legal certainty has always been an overarching factor in deciding these sorts of issues.

Even so, there have been instances where despite a failure to challenge a decision in the correct way, the courts have pronounced on its validity. The most prominent exception to the Oudekraal rule is the collateral challenge, where a party raises the invalidity of a decision as a defence in proceedings where it is sought to be enforced. Such a challenge is ‘collateral’ because it is raised in proceedings not designed to impugn the decision in question. In a sense a collateral challenge can be considered a shield against the enforcement of invalid decision. Daniel Freund and Alistair Price, who provide useful commentary on Oudekraal and its application by the Constitutional Court, explain instances where collateral challenges of these sorts have been allowed:

Where an administrative act is advanced as the legal justification for coercing a subject, [that is, requiring certain conduct by threatening a sanction, such as criminal punishment or civil liability, for non-compliance] and that subject believes the act to be unlawful, he or she may choose to ignore it and if he or she is then criminally prosecuted or sued in civil proceedings, the subject is entitled to defend herself by reactively challenging the administrative act in defence.

As an example, they refer to S v Smit where an accused who had been charged for failing to pay tolls successfully raised the unlawful declaration of the toll road as his defence. If Smit was not allowed to challenge the decision collaterally, he could have been coerced on the basis of an unlawful decision. Another famous case is the decision of the UK House of Lords in Boddington where an accused charged with smoking on a train in breach of a by-law, was allowed to raise its alleged invalidity collaterally, even though the House of Lords ultimately found against the accused on the question of the by-law’s validity.

---

87 See Wolgroeiers Afslaers (Edms) Bpk v Munisipaliteit van Kaapstad 1978 (1) SA 13 (A); Promotion of Administrative Justice Act 3 of 2000 (‘PAJA’) s 7(1); and City of Cape Town v Aurecon South Africa (Pty) Ltd [2017] ZACC 5, 2017 (4) SA 223 (CC) at fn 30.
89 For a discussion regarding the discretionary nature of the courts’ power to set order aside, see G Quinot & P Maree ‘The Puzzle of Pronouncing on the Validity of Administrative Action on Review’ (2015) 7 Constitutional Court Review 27.
90 Wade & Forsyth (note 18 above) at 251, the same sentence from the 7th edition of the work (at 342–344) was cited with approval in Oudekraal (note 6 above) at para 28.
91 Merafong (note 6 above) at para 23.
92 Tasima (note 6 above) at fn 60–62 (Court seems to have recognised that a collateral challenge is a species of a broader category of ‘reactive challenges’ which could include a formal counter-application which is more akin to a sword attacking the validity of the decision head-on. Any challenge to a decision raised in response to an application for enforcement is reactive on this definition.)
93 Freund & Price (note 5 above) at 189.
94 S v Smit 2007 (2) SACR 335 (T).
95 Boddington v British Transport Police [1999] 2 AC 143.
accused faced the potential risk of imprisonment for failing to comply with a requirement that may not have existed in law.\textsuperscript{96} Not all administrative decisions are coercive in nature, and presumably a court might be less inclined to allow a collateral challenge where that is the case (although the jurisprudence indicates that collateral challenges of administrative decisions shouldn’t only be available against certain categories of errors or be available only to certain categories of litigants).\textsuperscript{97}

This all has quite big implications for invalid court orders, because they are always coercive in nature. A party who has a court order in hand can utilise the powers of the state to enforce it. An order \textit{ad factum praestandum} – to do or refrain from doing something – is capable of being enforced through contempt proceedings and the threat of imprisonment. In contrast, an order \textit{ad solvendam pecuniam} – one which sounds in money – is enforceable through execution against property.\textsuperscript{98}

In \textit{Matjhabeng}, Nkabinde ADCJ emphasised that being committed to prison for contempt violates the rights not to be deprived of freedom arbitrarily or without just cause and not to be detained without trial, and impacts upon fair trial rights.\textsuperscript{99} The \textit{Motala} exception recognises that just cause begins with authority in law, and that a person should not lose their freedom without that foundation. The Constitutional Court on numerous occasions has also held that executing an order which sounds in money has the potential to affect property, housing and fair trial rights.\textsuperscript{100}

While the \textit{Motala} exception would frustrate a judgment creditor from exercising her rights accruing under an invalid order, a judgment debtor too has a right not to be ‘deprived of property except in terms of law of general application’,\textsuperscript{101} which must include the courts’ powers of adjudication. After all, even if a judgment creditor were to have a valid cause of action in law, if they bring their case in a court that doesn’t have authority over their dispute, their right to enforce their cause of action should not be permitted to trump the rights of the debtor to have their case adjudicated by the correct judicial officer. The machinery of the State is still involved in ensuring compliance with money judgments: through execution by the Sheriff. The state would still be coercing compliance with something illegal.

\textsuperscript{96} That result would be in breach of the \textit{nulla poena sine lege} principle.
\textsuperscript{97} \textit{Merafong} (note 6 above) at para 81, where Cameron J discusses how the approach must be flexible and less categorical, overturning the Supreme Court of Appeal’s ruling in the case to the effect that public functionaries may not raise collateral challenges (\textit{Merafong City Local Municipality v AngloGold Ashanti Limited} [2015] ZASCA 85, 2016 (2) SA 176 (SCA) at para 17). See also Wade & Forsyth (note 18 above) at 235–238 (Describe instances where collateral challenges are permitted in the United Kingdom.)
\textsuperscript{98} \textit{Matjhabeng Local Municipality v Eskom Holdings Limited & Others; Mkhonto & Others v Compensation Solutions (Pty) [2017] ZACC 35, 2018 (1) SA 1 (CC) at paras 56–57. There appears to be one blend between them: the refusal to pay maintenance pursuant to a court order can result in contempt and committal to prison \textit{Bannatyne v Bannatyne} [2002] ZACC 31, 2003 (2) SA 363 (CC).}
\textsuperscript{99} \textit{Matjhabeng} (note 98 above) at para 1, citing Constitution ss 12(1) and 35(3).
\textsuperscript{100} Constitution ss 25(1) and 34. See \textit{University of Stellenbosch Legal Aid Clinic & Others v Minister of Justice & Correctional Services & Others; Association of Debt Recovery Agents NPC v University of Stellenbosch Legal Aid Clinic & Others; Matava Trading 279 (Pty) Ltd & Others v University of Stellenbosch Legal Aid Clinic & Others [2016] ZACC 32, 2016 (6) SA 596 (CC) at para 129.}
\textsuperscript{101} Constitution s 25(1).
In short, if the courts are willing to allow collateral challenges to coercive administrative decisions – and in effect treat them as nullities – then it is reasonable for courts to treat orders made without authority on a similar basis.102

There may be other reasons for the differentiation in treatment. First, judges occupy a different position from administrators and other decision-makers in our democracy. They are required to be ‘appropriately qualified’103 and part of that means being well versed in the law, a requirement that is not applicable to administrators. (I mean this as a general proposition; many people in the public administration are well qualified in the law, but that is not always a requirement for a position.)

Secondly, challenging a court order is a far stricter process than taking an administrative or executive decision on review. In civil matters, leave to appeal against a High Court order must be applied for within fifteen court days104 and, while condonation is possible for late applications, it is not readily given.105 An application for rescission of judgment must be brought within twenty days106 and the courts have held that a similar period would be considered reasonable for the purposes of rescission applications brought in terms of the common law.107 In contrast, in respect of administrative action, the time frames are more lenient. PAJA specifies that applications for review must be brought within a reasonable time and not more than 180 days of the aggrieved party having become, or ought reasonably to have become, aware of the impugned decision, or after internal remedies are exhausted.108 The time period can be extended by agreement or by the court if the interests of justice require.109 There has been at least one case where a fifty year delay was condoned – the second instalment of the Oudekraal matter.110 Under the principle of legality, review challenges must also be brought within a reasonable time, although with no set period,111 and unreasonable lateness can be condoned by

---

102 One of the anonymous reviewers made the point that some of the concerns raised in the article about coercion possibly could be dealt with in other ways: committal to prison for contempt requires proof beyond a reasonable doubt that a party willfully and in bad faith ignored an order (Matjhabeng (note 98 above) at para 103), therefore a party confronted with allegations of not complying with an invalid order might defend itself by raising its belief that it was not bound by the order due to its invalidity. The main problem with this is that for civil debts there is no enquiry as to why the judgment debtor ignored a court’s order; a warrant of execution is simply issued and executed and so those safeguards are not present.

103 Constitution s 174(1) (‘Any appropriately qualified woman or man who is a fit and proper person may be appointed as a judicial officer. Any person to be appointed to the Constitutional Court must also be a South African citizen.’)

104 Rules regulating the conduct of the proceedings of the several provincial and local divisions of the Supreme Court of South Africa Government Notice R 48 (12 January 1965) rule 49(1)(b). That is within 15 days of the order or the court’s reasons being handed down.

105 See the discussion in van Loggerenberg (note 75 above) at D1-663 and D1-669. See also Van Wyk v Unitas Hospital & Another [2007] ZACC 24, 2008 (2) SA 472 (CC).

106 Rule 31(2)(b).


108 PAJA s 7(1).

109 PAJA s 9(1).

110 Oudekraal Estates (Pty) Ltd v the City of Cape Town & Others [2009] ZASCA 85, 2010 (1) SA 333 (SCA).

111 Aurecon (note 87 above) at fn 30.
a court, as it was in *Tasima*.\(^{112}\) Just recently, the Constitutional Court has held that in legality self-review cases of ‘clearly and indisputably unlawful’ conduct, it might still be required to look into the merits of the case even where it finds that delay should not be overlooked.\(^{113}\) Moreover, collateral challenges of administrative decisions are not usually time-barred.\(^{114}\)

Thirdly, as has been remarked by many over the centuries,\(^{115}\) the judiciary is the weakest branch of state and having no army or sword of its own to enforce its orders, it must rely on moral authority in society to fulfil its mandate as the interpreter of the Constitution and upholder of the law.\(^{116}\) If the judiciary – as the ultimate interpreter of the law – makes mistakes as to its adjudicative powers, and the same branch then enforces those mistakes through contempt or execution, that might undermine the vital moral authority the judiciary requires to play its part in society. Because judges decide on the legality of all public conduct, society needs to trust that judges will not abuse their authority when they do so.

IV WHY TREAT AUTHORITY-ERRORS DIFFERENTLY FROM OTHER ERRORS IN RESPECT OF COURT ORDERS?

That may be so, but it still doesn’t answer the question why the courts draw a distinction between court orders made without authority and those issued where another error was committed. All irregular orders are coercive, no matter the nature of the irregularity.

Notably, the Constitutional Court has refused to make the same distinction in respect of administrative and executive decisions, despite some uncertainty in the jurisprudence. In a 2009 article commenting on *Oudekraal*, Daniel Pretorius took the view that the law continued to draw such a distinction, by explaining: ‘An act which is ultra vires (in the sense that the decision-maker had no jurisdiction to perform that act) may be ignored by its author, without any need for revocation or judicial review.’\(^{117}\) In *Kirland*, Cameron J seemed to imply the same by holding: ‘despite its vulnerability to challenge, [it was] a decision taken by the incumbent of the office *empowered to take it*, and remained effectual until properly set aside.’\(^{118}\) In *Merafong* the same Justice firmly rejected this approach. Counsel for one of the parties, relying on the quoted sentence from *Kirland*, argued that where a decision ‘lacks the facial imprimatur of lawfulness’ it is a nullity and can be ignored.\(^{119}\) Cameron J disagreed, pointing out that administrative decisions made without authority – or ‘not authorised by an empowering provision’ – are subject to review under PAJA,\(^{120}\) and that ‘[a] decision taken within statutory

\(^{112}\) *Tasima* (note 6 above) at paras 151–152. Notably the Court avoided specifying which regime, PAJA or under the principle of legality, it was utilising to review the extension of the agreement. In light of the decision in *State Information Technology Agency Soc Ltd v Gijima Holdings (Pty) Ltd* [2017] ZACC 40, 2018 (2) SA 23 (CC) it was probably legality review.

\(^{113}\) *Buffalo City Metropolitan Municipality v Asla Construction (Pty) Limited* [2019] ZACC 15, 2019 (4) SA 331 (CC) at para 66.

\(^{114}\) Freund & Price (note 5 above) at 190.

\(^{115}\) Montesquieu *The Spirit of the Laws* Vol I (1777) 204; Hamilton *The Federalist No. 78* (1788); and A Bickel *The Least Dangerous Branch: The Supreme Court at the Bar of Politics* (1962).

\(^{116}\) *S v Mamabolo* [2001] ZACC 17, 2001 (3) SA 409 (CC) at para 16.

\(^{117}\) Pretorius (note 11 above) at 548.

\(^{118}\) *Kirland* (note 6 above) at para 105 (emphasis added.)

\(^{119}\) *Merafong* (note 6 above) at para 50 and fn 69.

\(^{120}\) Ibid at paras 50–52.
authority may be equally plainly vitiated – for instance by palpable fraud, or error of law, or mistake of fact.”\textsuperscript{121}

The line of argument the Court rejected is very similar to a distinction drawn between jurisdictional errors of law and non-jurisdictional errors of law in pre-1994 South African administrative law.\textsuperscript{122} Under earlier authority, like Chesterfield House,\textsuperscript{123} the courts would not interfere with errors of law committed by administrators when they were acting within their jurisdiction and where that caused no consequential irregularity. These were errors made on the merits; they were ‘non-jurisdictional errors’, which were ‘regrettable but not reviewable.’\textsuperscript{124} Only those errors which caused an administrator to fail to appreciate its powers or to misconstrue them were reviewable – ‘jurisdictional errors’.\textsuperscript{125} In Hira, Corbett CJ made the artificiality of this distinction clear, explaining that any error of law, even one made within a functionary’s jurisdiction, has the potential to cause an administrator to misconstrue its powers.\textsuperscript{126} Under the constitutional era, where the judiciary has the ultimate authority to decide questions of law, the Constitutional Court has rendered this distinction obsolete.\textsuperscript{127}

Even so, treating all unlawful government decisions as binding and effective, irrespective of whether they are made with or without authority, does produce some bizarre results. There may, for instance, be a situation where a functionary purports to make a decision it so obviously has no authority in law to take.\textsuperscript{128} Picture an ombudsperson who, exercising her powers to take appropriate remedial action, instructs Parliament to amend the Constitution because she believes that the founding law is standing in the way of resolving a matter she has investigated. Parliament, undoubtedly, is the only functionary empowered to make amendments to the Constitution,\textsuperscript{129} and the ombudsperson’s authority cannot extend that far. For the sake of

\textsuperscript{121} Ibid at para 53.
\textsuperscript{122} Hoexter (note 11 above) at 282–290 and Baxter (note 17 above) 468–472.
\textsuperscript{123} Johannesburg City Council v Chesterfield House (Pty) Ltd 1952 (3) SA 809 (A) (‘Chesterfield House’).
\textsuperscript{124} Hira & Another v Booyens & Another [1992] ZASCA 112, 1992 (4) SA 69 (A) (‘Hira’) at 84A.
\textsuperscript{125} Local Road Transportation Board & Another v Durban City Council & Another 1965 (1) SA 586 (A) (Administrator refused to renew certain certificates on the ground that they didn’t exist and that there were thus no certificates before it. Holmes JA held that the decision was erroneous in law and that this precluded the administrator from considering the applications for renewal. It was thus a reviewable error.)
\textsuperscript{126} Hira (note 124 above) at 90F. The Appellate Division did not completely obliterate the distinction, but held that jurisdictional errors of law will be reviewable, unless it was the intention of the legislature, gleaned from the administrator’s founding statute, that the administrator would exercise the exclusive authority to decide the question of law concerned. Ibid at 93C. The UK House of Lords also (mostly) abandoned this distinction in a number of judgments in the latter half of the twentieth century. Anisminic Ltd. v Foreign Compensation Commission [1969] 2 AC 147; O’Reilly v Mackman [1983] 2 AC 237; Reg v Hull University Visitor, Ex parte Page [1993] AC 682; and Boddington (note 95 above).
\textsuperscript{128} Freund & Price (note 5 above) at 188 (‘An attempt by the National Heritage Council to dismiss the National Director of Public Prosecutions, to take an improbable example, is surely null and void.’)
\textsuperscript{129} Constitution s 44(1)(a)(i) and (b)(i).
argument, let’s imagine that the decision is never taken on review and the ombudsperson brings judicial proceedings to force Parliament to comply with the instruction. It is very unlikely that a court would enforce the instruction even though it hadn’t been challenged. The likely scenario is that the court would regard the instruction as a nullity for having no basis in law: it would be wary of ordering an amendment to the Constitution in this manner. This is only partly a hypothetical situation, because a similar order made by the Public Protector was indeed taken on review and set aside. But it is an example of a more obvious case of a decision of a State institution that probably should not be treated as effective (ie, another exception to the Oudekraal rule).

Returning to invalid court orders, curiously it was Cameron J who reaffirmed the Motala exception in Kirland. Buried in a footnote, he made the following remark in passing:

[Motala provides] that judicial decisions issued without jurisdiction … are nullities that a later court may refuse to enforce (without the need for a formal setting aside by a court of equal standing). This seems paradoxical but is not. The court, as the fount of legality, has the means itself to assert the dividing line between what is lawful and not lawful. For the court itself to disclaim a preceding court order that is a nullity therefore does not risk disorder or self-help.

131 Other than that, the courts have not given further reasons as to why they draw a distinction between authority-related errors and other types in respect of court orders. After all, if a party were seeking to enforce an order which was made unlawfully due to another error and not challenged, the court enquiring into its invalidity would still be asserting the ‘dividing line’ between what is and what is not lawful. And a person may still be deprived of their freedom or property through coercion by unlawful government conduct.

Admittedly, this is the weakest part of the reasoning in the jurisprudence recognising the Motala exception. It is also difficult to reconcile with the remarks made by Khampepe J in Tasima that all court orders exist in fact and have consequences in the real world (much the same as administrative decisions), that judges also err sometimes, and that a culture of disobeying court orders could have negative effects for the rule of law. A court hearing an appeal also has the power to overturn an order made without authority. In short, the same ‘rule of law reasons’ regarding certainty are present no matter the nature of the error.

Perhaps the courts have been making the following distinction. Some irregular court orders are void where they are made without authority in the sense described in this article. Other irregular court orders are simply voidable since they are vitiated for non-authority related reasons. The courts treat the former as nullities and invalid, and the latter as binding, unless they are properly challenged and set aside, ie, made void. But this is unlikely, as drawing the distinction contradicts the general proposition of legality in our public law that all unlawful acts are invalid and void, and are not made void by judicial pronouncement. Lawrence Baxter once noted that the law does not treat undetected crimes as lawful – why then should an undetected

130 South African Reserve Bank v Public Protector & Others [2017] ZAGPPHC 443, 2017 (6) SA 198 (GP) at paras 42–44. Murphy J didn’t exactly take the same line as I have here. He found that the type of order made was not within the scope of the investigation concerned and was thus invalid. He also took issue with the order breaching the separation of powers.

131 Kirland (note 6 above) at fn 78.

132 This distinction between void and voidable acts once featured in English administrative law, but was rejected by the House of Lords in Anisminic Ltd (note 126 above). It also sometimes featured in our law right up until the new constitutional dispensation. L. Rose Innes Judicial Review of Administrative Tribunals in South Africa (1963) 92; Coalcor (Cape) (Pty) Ltd v Boiler Efficiency Services CC 1990 (4) SA 349 (C)).
unlawful government act be treated as lawful because it has not been challenged? In short there are some conceptual and justification-related difficulties with the Motala exception.

It seems to me that the courts treat authority-related errors differently from others in respect of invalid court orders, simply because they can. What I mean is that first, a court deciding whether to enforce an extant order can in many instances determine the question of authority as an objective legal question. To do so the court need not have access to the entire record of the proceedings. It doesn’t even need to look any further than the paper the order itself is printed on. Motala is an example where the order concerned quite clearly contradicted an earlier decision on the point; all Ponnan JA had to do was hold it up to the law. Other kinds of non-authority related errors are more difficult to ascertain: they cannot be considered simply with reference to the law alone and thus often require a fuller consideration of the facts and merits of the case. An appeals process is ideally suited to making those sorts of determinations. Secondly, the common law has always provided a rule which allowed the courts to do so, and so the courts do it because they can. ‘This may seem unsatisfactory to some, but as Froneman J remarked in Bengwenyama ‘the law often is a pragmatic blend of logic and experience.’

What Kirland and Merafong do not explain is why the same exception should not be recognised in respect of administrative decisions. An organ of state can equally ignore both its own decision or an order of a court, and as I pointed out above, there may be instances of a patently invalid administrative decision for lack of authority. The reason seems to be that the invalid government conduct comes from the judiciary itself and the judges are not comfortable with enforcing authority-related mistakes of their own. Legality here outweighs certainty on the spectrum of the rule of law.

This problem came up in another recent decision of the Constitutional Court, ACSA. In the case a tender had been set aside by the High Court in review proceedings. One of the tenderers lodged an appeal against the High Court order, but the decision-maker didn’t participate in its hearing and abided by the appeal court’s decision. After the appeal hearing but before judgment was handed down, the two tenderers found common ground and in a settlement agreed to set aside the High Court order with the intention of reviving the tender as it had been originally awarded. The settlement agreement was made an order of court by the full court. When faced with this order, the decision-maker refused to take steps to give effect to the tender award. The successful tenderer thus approached the High Court to enforce the court order.

Hughes J in the High Court dismissed the enforcement application. She reasoned that the original High Court order was an order in rem that affects not only the parties to the dispute but the world at large – a ‘public remedy’. As such, an agreement between private parties could not set it aside. The full court order in the settlement agreement therefore did not bind her.

133 Baxter (note 17 above) at 360.
134 This highlights the importance of stare decisis, which I discuss below in part VI.C.
135 See discussion about Voet’s views above in part II.B.
136 Bengwenyama Minerals (Pty) Ltd (note 6 above) at para 85.
137 Airports Company South Africa v Big Five Duty Free (Pty) Limited & Others [2018] ZACC 33, 2019 (2) BCLR 165 (CC) (‘ACSA’).
138 Ibid at para 7.
The Supreme Court of Appeal rejected this reasoning. While assuming that an order setting aside an administrative decision is a ‘public remedy’,140 Lewis JA focused on interpreting the settlement agreement in light of its purpose and the parties’ intentions. The decisive paragraph in her judgment reads:

What was the purpose of the withdrawal and abandonment, coupled with the agreement of settlement, if not to set aside the [original high court order]? It could be none other than to agree that the award … was to stand. There is no other purpose that the parties could have intended to achieve.141

The Constitutional Court, on appeal, produced three judgments with very different reasoning. Froneman J for the majority142 began his decision by stating two propositions:

The first is that a judgment \textit{in rem} may not be set aside by only a settlement agreement between the litigating parties in an appeal against that judgment. For a judgment \textit{in rem} to be set aside by a settlement agreement, the court hearing the appeal must give its sanction to the agreement being made an order of court on the basis that the setting aside is justified by the merits of the appeal. \textit{The second is that the court sanctioning the settlement agreement should give its reasons for doing so}.143

On this approach, the courts indeed have authority to set aside orders \textit{in rem} but must do so in a certain way. Froneman J held that the original High Court order was an order \textit{in rem} and ‘withdraw’ could not mean that it was set aside in the absence of an indication from the full court to that effect.144 Froneman J further criticised the reasoning of the Supreme Court of Appeal, that no matter the parties’ true intentions, the settlement agreement could not have had the effect of setting aside the High Court order.

Jafta J concurred with the majority but for very different reasons. He held that the settlement agreement should never have been made as its terms were ‘inconsistent with the Constitution and this rendered the order of the full court invalid.’145 He placed emphasis on the tender’s invalidity for its inconsistency with s 217 of the Constitution. Citing \textit{Eke}, Jafta J held that the settlement agreement did not meet the requirements for it to be made an order of court.146

Cachalia AJ dissented. Placing direct reliance on \textit{Eke}, \textit{Tsoga Developers} and \textit{Tasima} he held:

Once a settlement agreement is made an order of court, it has the same standing and qualities as any other court order. Its effect is to finally determine the status of the rights and obligations between the parties and the issues covered by the dispute. It is enforceable like ‘any other’ court order. This means that it may only be impugned through a ‘legally cognisable’ process such as rescission or appeal.147

He held that the full court order, like that of the High Court, was also an order \textit{in rem} as it finally determined the latter’s status and the administrative decision underlying it, for all the parties to the case – including the decision-maker. Cachalia AJ cited \textit{Tasima} for the proposition that the decision-maker could not ignore the full court order.148 Importantly, the

\begin{enumerate}
\item Ibid at para 26.
\item ACSA (note 137 above) at para 1 (emphasis added.)
\item Ibid at para 45.
\item Ibid at para 82.
\item Ibid at para 81.
\item Ibid at para 93, citing \textit{Eke} (note 37 above) at paras 29 and 31, and \textit{Tsoga Developers} (note 15 above) at para 52.
\item ACSA (note 137 above) at para 96.
\end{enumerate}
full court indeed had the authority in law to issue the order it did incorporating the settlement agreement. Cachalia AJ also stressed that the full court made the settlement agreement an order of court after hearing the case, which ‘necessarily implies that the full court considered the merits of the dispute’ as required by the first of Froneman J’s two propositions.149 Turning to interpret the agreement, Cachalia AJ disagreed with the majority and held that what the parties sought to do was set aside the High Court order, and since the full court had the authority to do so in the circumstances, that should be the meaning ascribed to the agreement.150

Analysing the case using the Motala exception, we might ask whether the full court had the authority to overturn an order in rem? According to Froneman J’s judgment the answer is ‘Yes’, if it decided the merits and gave reasons for doing so. If that is so, then the Motala exception cannot apply in these circumstances, ie, the Full Bench had authority to do what it did, the issue was whether it exercised its authority appropriately. In determining whether it made a valid decision, a later court would have to look further than the full court order itself, in that it would need to see whether that order was backed up by reasons and whether the court looked into the merits. Therefore, in the absence of an appeal against the settlement agreement made an order by the full court, it is binding.

This means that the approach of the majority is open to some doubt since it sought out to determine whether the full court had given reasons for making the settlement agreement an order when the original order was not on appeal before the Constitutional Court. The irregularity did not simply jump from the page. Jafta J’s judgment is problematic because he too would have to look at more than just the order itself to determine whether it was made unlawfully; his judgment would extend the Motala exception beyond its narrow application. His reliance on Eke is misplaced, for as I pointed out above, Tsoga Developers clarified that Eke dealt with the requirements a court should be satisfied of before making a settlement agreement an order, and reiterated that once made it is binding like any other.151 On both of these approaches, this appears to be a case in which the Constitutional Court was looking to justify a perceived better outcome (ie, the tender being set aside) and relaxed the rules pertaining to the binding nature of the order which did not support that outcome.152 That leaves Cachalia AJ, whose judgment best applies the general rule that unchallenged court orders bind, other than the narrow exception carved out in Motala. The exception was not applicable in the case because the full court had authority to sanction the settlement agreement.

V Why the courts might still enforce invalid orders

At this point the reader may be concerned that the Motala exception goes too far and could encourage bad behaviour, particularly by the state. In unclear cases, public officials might ignore court orders instead of challenging them on appeal, wait for a private party to haul them before court in contempt proceedings and only then raise the order’s invalidity as their defence. That could open the floodgates of recalcitrant behaviour, undermine the administration of justice and jeopardise certainty.

149 Ibid at para 98.
150 Ibid at paras 109–110.
151 Tsoga (note 15 above) at para 52.
According to the Constitutional Court, a court may still elect not to disregard an invalid order using the *Motala* exception and proceed to determine whether there was contempt of it. In *Tasima*, Khampepe J held that contemptuous conduct does not only concern the order itself but also ‘violat[es] the dignity, repute or authority of the court.’\(^{153}\) She explained:

> [W]hile a court may, in the correct circumstances, find an underlying court order null and void and set it aside, this finding does not undermine the principle that damage is done to courts and the rule of law when an order is disobeyed. A conclusion that an order is invalid does not prevent a court from redressing the injury wrought by disobeying that order, and deterring future litigants from doing the same, by holding the disobedient party in contempt.\(^{154}\)

The objective of contempt is to maintain the rule of law, rather than to punish a party who breaches an order.\(^{155}\) There may be instances where a court order is disregarded with such malice, instead of being challenged on appeal or in a rescission application, that a court finds the conduct to be contemptuous.

But this should be seen as the exception to the *Motala* exception. There is a stronger case where it is a public functionary or organ of state that ignores an order. Section 165(4) of the Constitution specifies that ‘[o]rgans of state, through legislative and other measures, must assist and protect the courts to ensure the independence, impartiality, dignity, accessibility and effectiveness of the court.’ Khampepe J pointed out in *Tasima* that there is a higher duty for organs of state to take orders they disagree with on appeal.\(^{156}\) That is what is required of organs of State as ‘good constitutional citizens’.\(^{157}\) The state has the resources to do so after all.\(^{158}\)

In any event, the state may not be able to comply with every court order or challenge it. The facts in *Motala* are again instructive in this regard. The Master’s Office had previously approached the High Court for a declaratory order that it alone is vested with the competence in law to appoint judicial managers. This was done in response to an emerging trend of court ordered appointments of judicial managers, which was said to undermine the Master’s ability to carry out its legislative duties properly.\(^{159}\) Despite the declaratory order, and a detailed judgment setting out the basis for such conclusions by Bertelsmann J, judges in the same Division of the High Court continued to issue such orders, presumably *ex tempore* and in busy motion court after receiving draft orders from counsel, and with no reasons given. While the Master’s Office is cited in every insolvency proceeding instituted in court, can we realistically expect it to instruct lawyers and appear in every matter to ensure that court orders of this kind are not granted, or to appeal the orders that have been issued? Does the Master have the time and resources to do so?\(^{160}\)

For private parties an even stronger case of contempt and bad faith would have to be made out because they have no additional obligation to assist the courts in enforcing orders.

---


\(^{154}\) *Tasima* (note 6 above) at para 186.

\(^{155}\) *Matjhabeng* (note 98 above) at para 57 citing *Mjeni v Minister of Health and Welfare, Eastern Cape* 2000 (4) SA 446 (Tk) at 456B–C.

\(^{156}\) *Tasima* (note 6 above) at para 187.

\(^{157}\) *Merafong* (note 6 above) at paras 59–61.

\(^{158}\) *Tasima* (note 6 above) at para 187 and *Meadow Glen Home Owners Association v City of Tshwane Metropolitan Municipality* [2014] ZASCA 209, 2015 (2) SA 413 (SCA) at para 8.

\(^{159}\) *Ex Parte Master of the High Court* (note 31 above) at para 39.

\(^{160}\) *Motala v The Master of the North Gauteng High Court, Pretoria* [2019] ZASCA 60 at para 2 (Supreme Court of Appeal noted that the Pretoria Master’s Office deals with approximately 7000 insolvent estates every year and receives around 750 liquidated companies to administer per year.)
VI HOW TO DETERMINE JUDGES’ ADJUDICATIVE POWERS

Having explained why it is necessary to recognise the Motala exception, why authority-errors should be treated differently from others and why there may be an exception to the Motala exception, I now turn to the issue of ‘how’ the courts are to determine whether an order was issued without authority.

The courts’ adjudicative powers delimit their functions in two ways, first in respect of what cases they can hear and decide (subject-matter authority) and secondly in respect of the kinds of orders they can make (court-order authority). To determine whether an order is invalid, one must ask objectively whether the court possessed the authority to make the order. Another court will be competent to make that determination since it only concerns an objective legal question divorced from the particular facts in the matter. This also means that whether a court correctly asserted jurisdiction over a matter on a more general basis – for example the appropriateness of an order allowing the attachment of property to found jurisdiction or whether a party is domiciled in a court’s area of jurisdiction – must be determined in the normal way, namely by appeal or rescission.161

A Subject-matter authority

The types of matters that may be decided by all courts within the judicial hierarchy are delineated either in the Constitution, legislation,162 the common law or customary law. The Constitutional Court may only decide constitutional matters and arguable points of law of general public importance and has exclusive jurisdiction over certain issues;163 the Supreme Court of Appeal may only decide appeals from the High Court, and no competition or labour matters;164 and the High Court may decide constitutional matters unless the Constitutional Court has elected to hear the matter directly or where the matter has been assigned to another court of similar status by legislation,165 and any other matter not assigned in an Act to another court.166 The superior courts are also granted inherent jurisdiction.167

Section 21 of the Superior Courts Act 10 of 2013 specifies that the High Court ‘has jurisdiction over all persons residing or being in, and in relation to all causes arising and all offences triable within, its area of jurisdiction and all other matters of which it may according to law take cognisance.’ The commentary in Erasmus suggests that ‘jurisdiction’ used here refers to its power to ‘hear, adjudicate upon, determine and dispose of disputes between parties in a matter brought before it.’168 That accords with the definition of ‘authority’ used in this article.

Section 170 of the Constitution empowers Parliament to determine matters to be decided by other courts. The District and Regional Magistrates Courts – those famous creatures of

161 Compare Tralex Limited v Maloney & Another [2016] ZASCA 128 at para 16. The Supreme Court of Appeal seemed to indicate that ‘lack of jurisdiction’ for the purposes of the Motala exception can be understood more widely than relating merely to the court’s authority, such as whether a court should have ordered an attachment to found or confirm jurisdiction. This ruling appears to be inconsistent with the rest of the jurisprudence.

162 Constitution s 170 (‘All courts other than those referred to in s 167, 168 and 169 may decide any matter determined by an Act of Parliament.’)

163 Constitution s 167(3) and (4).

164 Constitution s 168(3).

165 Constitution s 169(1)(a).

166 Constitution s 169(1)(b).

167 Constitution s 173.

168 Van Loggerenberg (note 75 above) at A2-89.
statute in our administration of justice – can only decide matters they are empowered to in legislation, primarily the Magistrates Courts Act 32 of 1944. Magistrates may also ‘not enquire into or rule on the constitutionality of any legislation or any conduct of the President.’

Even within their ranks, the Regional Courts are empowered to hear a broader range of both criminal and civil matters. For example, only a Regional Court may hear cases pertaining to the dissolution of marriages.

Parliament has created many specialised judicial tribunals to decide specific matters, including the Labour Court, Labour Appeal Court, Competition Appeal Court, and Land Claims Court. The authority of the High Court in these matters has in turn been denuded. And the Supreme Court of Appeal was also divested of some authority. If one of these specialist legal tribunals decides a matter outside of its authority, an order it issues would be invalid, and a High Court or the Supreme Court of Appeal deciding an issue reserved for a specialist tribunal would suffer the same deficiency. For example, the Land Claims Court has no power to decide matters under the PIE Act. Therefore, an order declaring occupants of land which is the subject of a land claim dispute to be ‘illegal occupiers’ for the purpose of PIE would be invalid.

B Court-order authority

There are other limits to the adjudicative powers of judges: the law at times does not grant a court the power to make an order, even in a matter in which it has subject-matter authority. In effect, while superior courts all have ‘inherent jurisdiction’, their powers to issue orders are, to varying degrees, circumscribed by law.

In the criminal realm for example, a judge cannot utilise their inherent jurisdiction to create new offences and sentences – that would be in breach of the right not to be convicted of a crime that was not an offence at the time. In S v Mhlakaza, after explaining that sentencing powers are derived primarily from statute, Harms JA held that the courts must ‘limit themselves to performing their duties within the scope of that jurisdiction.’ An order fixing a non-parole period for a time greater than two thirds of the sentence – the maximum
non-parole period allowed under the relevant statute — would therefore be invalid since it
would be made beyond the authority of the court. This is an example of how an invalid order
could have a deleterious effect on a person’s life if it were to be enforced.

Motala and Changing Tides are also good examples of instances where a court made an order
without authority. No one had argued that the High Court does not have jurisdiction to hear
insolvency proceedings or eviction matters. Rather the concern was that the court order issued
in each case was not competent under the relevant statute. Another example comes from a
judgment reported in the same volume of the South African Law Reports as Tasima, where a
High Court made a declaratory order that Magistrates have no power under s 87(1) of the
National Credit Act 34 of 2005 to vary the agreed interest rate in a credit agreement and that
any order to such effect is invalid. Many of these examples pertain not to organs of state
ignoring orders, but to ordinary people who may be affected by something that should not
have been ordered in the first place.

The superior courts also have powers to grant ‘appropriate relief’ and any order which is
‘just and equitable’ in constitutional matters – particularly those where rights are infringed
or threatened. Following declarations of invalidity of legislation, the Constitutional Court
has ordered a range of different remedies to regulate their impact including: limiting
the retrospective effect of an order to avoid the consequences flowing from its objective
invalidity; suspending the declaration in order to give Parliament time to rectify the defect and
extending suspensions where Parliament hasn’t acted timeously; severing words from
provisions which offend the Constitution if possible; and reading-in words to remedy an
omission that makes a provision unconstitutional. The Court has also ordered declaratory
relief and issued structural interdicts – particularly in socio-economic rights cases. And

---

181 Criminal Procedure Act 51 of 1977 s 276B(1):
   ‘(a) If a court sentences a person convicted of an offence to imprisonment for a period of two years or longer,
       the court may as part of the sentence, fix a period during which the person shall not be placed on parole.
   (b) Such period shall be referred to as the non-parole-period, and may not exceed two thirds of the term of
       imprisonment imposed or 25 years, whichever is the shorter.’

182 Nedbank Ltd v Jones [2016] ZAWCHC 139, 2017 (2) SA 473 (WCC).
183 Constitution s 38.
184 Constitution s 172(1)(b).
186 Executive Council of the Western Cape Legislature & Others v President of the Republic of South Africa & Others
   (CC) at para 50.
   SA 524 (CC).
189 Acting Speaker of the National Assembly v Teddy Bear Clinic for Abused Children & Another [2015] ZACC 16,
   2015 (10) BCLR 1129 (CC).
191 For a recent example see University of Stellenbosch Legal Aid Clinic (note 100 above).
193 Minister of Health & Others v Treatment Action Campaign & Others (No 2) [2002] ZACC 15, 2002 (5) SA 721
   (CC), and Pheko & Others v Ekurhuleni Metropolitan Municipality [2011] ZACC 34, 2012 (2) SA 598 (CC).
there is a fledgling jurisprudence relating to constitutional damages. These types of remedial orders have been carefully developed and justified by the Constitutional Court over the past two decades.

C Precedent as a tool

A useful tool in our law to determine whether a judge has authority is precedent or the *stare decisis* doctrine. It requires a judge to follow the decisions of courts higher up in the judicial hierarchy, whether it be an appellate court or a full court (three judges) or full bench (two judges) of their division, and also a single judge on the same level in their division. This stems from the need for certainty, stability and predictability in our legal system: ‘without [it,] deciding legal issues would be directionless and hazardous.’ Precedent is an important aspect of the rule of law in South Africa.

Only the *ratio decidendi* of a decision is binding – ie, the ‘[r]ationale or basis of deciding’. *Obiter dicta*, ‘things said by the way or in passing by a court … are not pivotal to the determination of the issue or issues at hand and are not binding precedent.’ Even so, the latter are persuasive, especially those which emanate from courts higher up in the hierarchy. Precedent may only be departed from where a court considers a judgment on the same level to be not simply wrong, but clearly wrong. A High Court for example, could not depart from a precedent set by the Supreme Court of Appeal even if the High Court regards it as clearly wrong. In a similar manner, the Supreme Court of Appeal cannot depart from a precedent set by the Constitutional Court.

For court order authority it works both ways. The jurisprudence may have held that a court has authority to hear a dispute and issue an order, and if so such orders will not be invalid due to an authority-related error. Or a precedent may have established that the court does not have the authority, with the concomitant result. Either way, a court is bound by precedent and if a judge issues an order, that an earlier judgment held the court has no authority to make, then one can easily ascertain that the order was invalid.

---

194 *Families of Mental Health Care Users Affected by the Gauteng Mental Marathon Project v National Minister of Health of the Republic of South Africa & Others* (Arbitration Award, 19 March 2018); *Komape & Others v Minister of Basic Education* [2018] ZALMPPHC 18; *Nqomane & Others v City of Johannesburg Metropolitan Municipality & Another* [2019] ZASCA 57 at paras 25–27.

195 *Turnbull-Jackson v Hibiscus Court Municipality & Others* [2014] ZACC 24, 2014 (6) SA 592 (CC) at paras 53–56; and *True Motives 84 (Pty) Ltd v Mahdi & Another* [2009], ZASCA 4, 2009 (4) SA 153 (SCA) at para 100. But see M Wallis ‘Whose Decisis Must We Stare?’ (2018) 135 South African Law Journal 1, 3 (Offers interesting remarks about how the Superior Courts Act may have altered the ‘traditional view’ concerning whether a judge is bound by decisions from judges in other divisions. Under the traditional approach, a decision from another division would only be persuasive, but now that may have changed and will be binding). For a discussion of the Constitutional Court’s relationship with precedent, see Brickhill, Corder, David & Marcus (note 10 above) at 2–12; J Brickhill ‘Precedent and the Constitutional Court’ (2010) 3 Constitutional Court Review 79.

196 *Turnbull-Jackson* (note 198 above) at para 53.

197 *Camps Bay Ratepayers’ & Residents’ Association* (note 6 above) at para 30.

198 *Turnbull-Jackson* (note 195 above) at para 61.

199 Ibid at para 57.

200 Ibid.
VII HARD CASES?

The focus of this article has been on the authority of High Court judges and Magistrates to hear cases and issues orders. The discussion has had regard mostly to cases of clear illegality, but there may be many instances where it is less clear that an order was issued without authority. There may very well be a penumbra of uncertainty as to whether a court has subject-matter or court-order authority in a case before it. For example, the question may never have been asked in a case before. And the discussion has not thus far taken account of how orders from the apex Court are to be treated.

A A spectrum of certainty

There are many obvious examples of instances where a court doesn’t have authority or competence to decide a case: the Supreme Court of Appeal could be hoodwinked into hearing a case as a court of first instance, any order it makes in those circumstances would be invalid and not binding; and a decree of divorce granted by a Magistrate sitting in the district court would too be invalid. These rules of adjudication are clearly laid down. But there may be many cases falling at the fringes still requiring judicial pronouncement.

A very recent example is *Greef*. The matter concerned whether Magistrates have the authority to interdict compliance with a subpoena issued during debt collection proceedings. There are two competing considerations. On the one hand, s 30 of the Magistrates’ Courts Act expressly provides that Magistrates have the authority to grant interdicts, subject to the ‘limits of jurisdiction prescribed by the Act’. On the other, Magistrates have no general authority to set aside subpoenas (they can only cancel them in limited instances and set aside improper service). Davis AJ held that an interdict against compliance with a subpoena would in substance amount to setting it aside, and concluded that such relief is ‘not legally competent’ in the Magistrates Court.

Another good example of uncertainty is the authority of the High Court to hear certain labour matters. Up until 2007, there was debate about whether an employee of the state could bring a labour dispute (such as a claim for unfair dismissal) before the High Court instead of the Labour Court, by framing their claim on the ground that their administrative rights had been breached. In *Chirwa*, a majority of the Constitutional Court held that such an applicant’s case, while being framed in administrative law was essentially a labour dispute under the LRA. Such cases should be brought in the Labour Court, which has exclusive jurisdiction, and not in the High Court. Cora Hoexter in her decisive analysis of the case, points out that the conclusion was more likely based on a policy reason – avoiding having

---

201 *Greef v Cooper & Others* [2018] ZAWCHC 170, 2019 (3) SA 203 (WCC).

202 Ibid at para 26, citing *S v Matisonn* 1981 (3) SA 302 (A) at 313E–F and *Marais v Smith* 2000 (2) SA 924 (W) at 933F–G.

203 *Greef* (note 201 above) at para 37.

204 I am indebted to one of the anonymous reviewers for referring this situation to me. For a very thorough discussion see Hoexter (note 11 above) at 210–218.

205 Section 157(2) of the LRA recognises concurrent jurisdiction between the High Court and the Labour Court in certain labour matters. *Fredericks & Others v MEC for Education and Training Eastern Cape & Others* [2001] ZACC 6, 2002 (2) SA 693 (CC) at para 41.

206 *Chirwa v Transnet Limited & Others* [2007] ZACC 23, 2008 (4) SA 367 (CC) (‘*Chirwa*’).

207 Ibid at para 65.
parallel systems of dispute resolution – rather than a legal one. Later in *Gcaba*, the Court attenuated this finding and seemed to recognise that where a decision has a wider impact than on the employee concerned, the High Court may have jurisdiction to decide the dispute based on the right to just administrative action.

An unpalatable consequence of the judgment in *Chriwa* in light of the *Motala* exception is that any labour matter, which was decided by the High Court during the decade between the LRA coming into effect and 2007, would be invalid and any order issued by the court would not be binding. That would have some very deleterious effects for legal certainty in South Africa.

Had this been raised by the parties, the Constitutional Court could have limited the retrospective effect of its new interpretation of LRA. For example in *Stratford*, the Court interpreted a provision of the Insolvency Act 24 of 1936 which requires service of an application for sequestration to be made on a person’s ‘employees’ as to include domestic employees. In doing so it overturned an earlier judgment of the Supreme Court of Appeal which held that the section did not require such service. The Court held that the declaration as to the correct interpretation of the provision should not be retrospective as many parties would have acted on the earlier decision of the Supreme Court of Appeal before that time, and that would unscrambled too many eggs. It ought to have done the same in *Chirwa*.

The question that needs to be answered however is this: should the existence of uncertainty surrounding whether courts have authority in a case count against the *Motala* exception to such an extent that the courts ought to abandon it in future? I would argue ‘no’. First, while law to many looks indeterminate, in truth on a day-to-day basis hard cases in the courts are unusual. Even lawyers closely associated with the American Legal Realism movement accepted this reality. Max Radin explained:

> ‘The law’ as a generalization of legal judgments is always incomplete since it is always concerned with a specific question not yet decided, as well as thousands already decided. The prognosis of that decision involves an estimate in advance of the factors that will determine the future judgment. In spite of the possible variety and number of these factors, the advance estimate is so highly probable in a number of cases that the statement of the law can be made with a fair degree of certainty and precision, and no decision will be required to test its accuracy since most men will regard the

---

208 Hoexter (note 11 above) at 213.
209 *Gcaba v Minister for Safety & Security* [2009] ZACC 26, 2010 (1) SA 238 (CC)(‘*Gcaba*’).
210 Ibid at para 66. Hoexter (note 11 above) at 213.
211 An issue that has also arisen in the UK. In the past, the courts held that they had no power to order that new understandings of legislative provisions (or their development of the common law) were to be prospective only. *Launchbury v Morgans* [1973] AC 127 (HL). But see *National Westminster Bank plc v Spectrum Plus Limited* [2005] 2 AC 680 (HL) at para 40 (House of Lords recognised that there may be some extraordinary instances where legal certainty requires that the court make its ruling prospective.)
212 *Stratford & Others v Investec Bank Limited & Others* [2014] ZACC 38, 2015 (3) SA 1 (CC)(‘*Stratford*’).
213 Ibid at para 37.
215 *Stratford* (note 212 above) at para 47.
216 This is not to say that the substantive justifications for various settled legal rules concerning adjudicative authority should not be questioned and tested, but rather that where a rule has been laid by the Constitutional Court or Supreme Court of Appeal in the past, those courts will be the appropriate forum to determine whether the rule is justified. See J Froneman ‘Legal Reasoning and Legal Culture: Our “Vision” of Law’ (2005) 16 *Stellenbosch Law Review* 3, 5–7 (Formal and substantive types of legal reasoning are explained and discussed).
decision as a foregone conclusion. Decisions will consequently be called for chiefly in what may be called marginal cases, in which prognosis is difficult and uncertain. It is this fact that makes the entire body of legal judgments seem less stable than it really is.\textsuperscript{217}

Secondly, there are tools in the law that the courts can use to ameliorate the sometimes-harsh consequences of the exception, such as limiting any finding that certain courts don’t have authority to apply prospective only. Thirdly, as I have stressed throughout this article, \textit{Motala} constrains the exercise of public power by requiring judges to decide cases only when they have authority to do so.

\textbf{B \textit{Motala} and the Constitutional Court}

Where \textit{Motala} finds little to no application is in respect of orders issued by the Constitutional Court, even if it is constrained by the law. As for its subject-matter authority, it is worth repeating that the Court may only hear two categories of cases: those which raise constitutional issues; and those which raise arguable points of law.\textsuperscript{218} That means that, strictly speaking, the Constitution does not grant the Court authority to decide cases that only raise factual disputes.\textsuperscript{219} This is not an unusual approach. Recently in \textit{Vedanta Resources}, the UK Supreme Court explained the point of limiting an apex court’s authority in this manner:

\begin{quote}
The essential business of this court is to deal with issues of law, rather than fact-finding or the re-exercise of discretion. The pursuit of detailed matters of factual (or evaluative) analysis in this court is therefore inappropriate, both because it is likely to involve a needless and useless misapplication of the parties’ time and resources, and because it distracts this court from its proper focus upon real issues of law.\textsuperscript{220}
\end{quote}

That said, the Constitution expressly grants the Constitutional Court the authority to make ‘the final decision whether a matter is within its jurisdiction’\textsuperscript{221} and, for this reason, it will be difficult if not impossible to argue that the Court ought not to have heard a case for want of authority.

Sometimes even the Justices of the Constitutional Court cannot find common ground on this issue: \textit{Jacobs}\textsuperscript{222} is one such case. It concerned the application of the common purpose doctrine in criminal law. The Constitutional Court split spectacularly about whether it had authority to hear the case. On one side, five Justices\textsuperscript{223} held that the court could not hear the matter because in essence it dealt with a factual dispute.\textsuperscript{224} On the other side, the remaining five Justices\textsuperscript{225} (the court sat as ten, an even number\textsuperscript{226}) held that the application of the common purpose doctrine was a constitutional issue and the Court had jurisdiction to decide the case.\textsuperscript{227}

\textsuperscript{217} Max Radin ‘In Defense of an Unsystematic Science of Law’ (1942) 51 \textit{Yale Law Journal} 1269, 1271 (emphasis added.)

\textsuperscript{218} Constitution s 167(3)(a) and (b).

\textsuperscript{219} Paulsen \& Another v Slip Knot Investments 777 (Pty) Limited [2015] ZACC 5, 2015 (3) SA 479 (CC) at para 20.

\textsuperscript{220} Vedanta Resources PLC \& Another v Lungowe \& Others [2019] UKSC 20 at para 12.

\textsuperscript{221} Constitution s 167(3)(c).

\textsuperscript{222} Jacobs (note 14 above).

\textsuperscript{223} Ibid (Goliath AJ; Cachalia AJ, Froneman J, Khampepe J and Madlanga J concurring).

\textsuperscript{224} Jacobs (note 14 above) at paras 46–51.

\textsuperscript{225} Ibid (Theron J; Zondo DCJ, Dlodlo AJ, Jafta J and Petse AJ concurring).

\textsuperscript{226} The Chief Justice was absent.

\textsuperscript{227} Jacobs (note 14 above) at paras 56 and 127.
In reaching this conclusion, the second group relied on *Makhubela v S*,228 a 2017 decision of the Court which seems to have held that the doctrine’s application raises a constitutional issue. In a concurring judgment aligned with the first five Justices, Froneman J229 held that the *Makhubela* decision had been made in error and should not be relied on.230

This was the first case in the Court’s history where it split evenly as to the outcome of a case and the order to be granted.231 Interestingly, while the Constitution explicitly allows the Court to sit in even numbers – since it specifies that at least eight Justices must sit to hear a case232 – it is silent as to what is to happen if the Court deadlocks. So, in keeping with the general thesis about the Constitutional Court’s authority discussed here, it was the Court that had to decide how to go forward. It concluded that in such instances the case should be dismissed, by holding: ‘There is thus no majority decision of this Court. The result is that the judgment and order of the full court stands.’233

The Court also has wide discretionary court-order authority when it decides constitutional matters. In *Corruption Watch*,234 Madlanga J contemplated just how wide this authority is, holding:

> In terms of section 172(1)(b) of the Constitution we may make any order that is just and equitable. The operative word ‘any’ is as wide as it sounds. Wide though this jurisdiction may be, it is not unbridled. It is bounded by the very two factors stipulated in the section – justice and equity. This Court has laid down certain principles in charting the path on the exercise of discretion to determine a just and equitable remedy.235

Even though it is wide (and justice and equity are often malleable) the Court has on occasion ruled that it cannot grant certain orders. In *Minister of Social Development*, the Court held that once an order suspending a declaration of invalidity of legislation has expired, it has no authority to grant an order reviving the suspension.236 In the case the suspension order concerned had expired on 5 March 2006 and the Minister’s application was only heard by the Court the next day. Van der Westhuizen J explained that while the Court has the power to extend a suspension before it expires, once the period has expired so too has the Court’s suspension power under s 172(1)(b)(ii). He further held that the Court has no authority to revive a law that is invalid, because that would in effect be legislating from the bench in breach of the separation of powers.237

So, arguably if the Court were to revive a suspension after its expiry, that would be an invalid court order. But that is not the end of the story. While the highest judicial organ is constrained by the law, it will ultimately be for the Court to decide whether it has authority to issue an order. It may in future come to the conclusion that in certain exceptional cases it indeed has authority to revive a suspension and hold that *Minister of Social Development* was...

---

228 *Makhubela v S; Matjike v S* [2017] ZACC 36, 2017 (12) BCLR 1510 (CC).
229 Ibid (Cachalia AJ and Madlanga J concurring.)
230 Jacobs (note 14 above) at para 97.
231 It has on occasion split evenly regarding the way in which the order was reached. *Myathaza v Johannesburg Metropolitan Bus Services (SOC) Limited t/a Metrobus & Others* [2016] ZACC 49, 2018 (1) SA 38 (CC); *Mogaila v Coca Cola Fortune (Pty) Limited* [2017] ZACC 6, 2018 (1) SA 82 (CC).
232 Constitution s 167(2).
233 Ibid at para 3.
234 Note 6 above.
235 Ibid para 68.
236 *Ex Parte Minister of Social Development & Others* [2006] ZACC 3, 2006 (4) SA 309 (CC) at para 38.
237 Ibid at para 39.
wrongly decided. It will be for the Court to decide. In short, given that the Constitutional Court is the sole arbiter of its own authority to hear cases and issue orders on grounds of justice and equity, it seems that the Motala exception can never really apply to South Africa’s apex Court.\textsuperscript{238}

\textbf{VIII WHY MOTALA MATTERS}

A judge’s decision can be of great consequence, often of more consequence than any Act of Parliament or administrative decision. It can be the difference between imprisonment and freedom, financial ruin and salvation, even between losing and winning an election. For these reasons, Ronald Dworkin once quipped that ‘[i]t matters how judges decide cases’.\textsuperscript{239} Motala matters because it reminds us that when judges exercise their great powers, they too are constrained by the law.

In this article I have traced where the Motala exception came from, tried to explain its application, and provided a justification for it based on principles emanating from the rule of law. I have also offered some guidance as to how our courts can determine whether an order is made with or without authority and considered the limits of the exception, particularly that it cannot apply to orders issued by our apex Court.

On a spectrum of the rule of law with the principles of legality and certainty occupying its poles, the Motala exception may seem to lean more towards the former and, some might argue, at the expense of the latter. But this is not necessarily the case. It promotes judges only acting when they are competent to do so, as well as more predictable and certain decision-making from the courts. Understanding the Motala exception as requiring judges to act both lawfully and predictably by following the rules and not deciding cases without authority, can ultimately be a good thing for the rule of law.

\textbf{IX POSTSCRIPT}

After the finalisation of the manuscript, the Constitutional Court handed down three important judgments that deal with various issues raised in this article.

\begin{enumerate}
\item \textbf{Court-order authority:} In \textit{Mwelase and Others v Director-General for the Department of Rural Development and Land Reform and Another},\textsuperscript{240} the Court confirmed the existence of a new special remedy for systemic rights violations. Cameron J held that in exceptional cases the courts have the authority to appoint a Special Master to oversee and monitor the implementation of the executive’s constitutional and legal obligations. Pointing out that the form of relief is novel in South African law,\textsuperscript{241} he grappled with whether it would be competent for the courts to have such power. He concluded that it wouldn’t breach the separation of powers\textsuperscript{242} or amount to judicial overreach, as the Supreme Court of Appeal had held, and stressed that the courts ‘intervene only when the evidence and arguments compel them to

\textsuperscript{238} This conclusion doesn’t mean that we should not question and debate the Court’s exercise of its authority in the future. We should accept that, ultimately, all orders of the Constitutional Court are essentially binding and cannot be ignored based on the Motala exception.

\textsuperscript{239} R Dworkin \textit{Law’s Empire} (1986) 1.

\textsuperscript{240} \textit{Mwelase & Others v Director-General for the Department of Rural Development and Land Reform & Another} [2019] ZACC 30.

\textsuperscript{241} Ibid at para 38.

\textsuperscript{242} Ibid at paras 46–49.
conclude that the Executive or the Legislature has done wrong, or has not done enough. And when the courts intervene, they do so with necessary trepidation.¹²⁴³ This is an important case, because it highlights that new remedies – and a development of court-order authority – may be established by the apex Court to deal with novel issues and with sufficient justification, especially in constitutional cases.

2. **Oudekraal rule:** In *Magnificent Mile Trading 30 (Pty) Limited v Charmaine Celliers NO*,²⁴⁴ the Court reexplored the justification for *Oudekraal*. Madlanga J for the majority,²⁴⁵ adopted the terminology – similar to this article – that *Oudekraal* and Kirland set forth a ‘rule’ that ‘says an unlawful administrative act exists in fact and may give rise to legal consequences for as long as it has not been set aside.’²⁴⁶ Jafta J²⁴⁷ again disagreed and held that the principle of legality does not allow the recognition or enforcement of invalid administrative acts and stated that the *Oudekraal* rule is not a true reflection of our law.²⁴⁸ ‘[T]o say an invalid action remains valid’ Jafta J held ‘defies logic’.²⁴⁹ In response, Madlanga J pointed out that the *Oudekraal* rule is all about the rule of law and its requirement of legal certainty:

> Imagine the spectre of organs of state and private persons ignoring or giving heed to administrative action based on their view of its validity. The administrative and legal chaos that would ensue from that state of affairs is unthinkable. Indeed, chaos and not law would rule.²⁵⁰

He held that because unlawful decisions exist in fact they may have legal consequences.²⁵¹ After careful analysis of the Court’s earlier jurisprudence, Madlanga J rejected that the rule defies ‘logic’, and reaffirmed that an unlawful decision is to be treated as valid and that ‘[t]reating the invalid act as valid does not invest it with legal validity.’²⁵²

3. **Motala exception:** Finally, in *Moodley v Kenmont School*²⁵³ the applicant had obtained a costs order against the respondent school. The school did not appeal the order nor pay it. When the applicant tried to levy execution against assets of the school, he was met with s 58A(4) of South African Schools Act 84 of 1996 which precludes a public school’s assets being attached. In response he sought and was granted an order declaring the provision unconstitutional in the High Court. Madlanga J for a unanimous Constitutional Court declined to confirm the order.²⁵⁴ The problem for the school was that the costs order remained extant. In rejecting an argument that the order should not be enforced, Madlanga J reiterated that court orders must be binding for the sake of the rule of law.²⁵⁵ Even though the *Motala* exception would not apply, as a competent court had issued the costs order and no steps had been taken to set

---

²⁴³ *Ibid* at para 53.
²⁴⁴ *Magnificent Mile Trading* (note 6 above).
²⁴⁷ *Jafta J* wrote a judgment concurring with the outcome, but disagreed about the reach of the *Oudekraal* rule.
²⁴⁸ *Ibid* at para 85.
²⁴⁹ *Ibid* at para 89.
²⁵⁰ *Ibid* at para 50.
²⁵¹ *Ibid* at para 51.
²⁵² *Ibid* at para 60. Emphasis in the original.
²⁵⁴ While accepting the provision limits the rights to equality and dignity, the Court held that it seeks to protect the right to basic education and that any limitation is reasonable and justifiable under s 36(1) in light of its purpose of avoiding adverse effects that could be caused by the attachment of school assets.
²⁵⁵ *Moodley* (note 253 above) at para 36.
it aside, Madlanga J again briefly discussed the exception (without pronouncing either way) and said the following:

Not even cases like *Changing Tides* and *Motala* which were referred to in this Court’s judgment in *Tsoga* suggest that persons – natural or juristic – or organs of state have an entitlement to ignore court orders based on their understanding of their lawfulness. According to *Changing Tides* and *Motala*, it is a court that declares an order previously granted and against which there is no appeal a nullity. In terms of s 165(5) persons and organs of state just must obey court orders whatever their view of them might be, subject, of course, to their exercise of the right of appeal.256

The upshot is that in more direct *obiter* remarks the Court has raised some doubt whether the *Motala* exception will withstand scrutiny before it, as on the approach set out a party must follow an order subject to the right to appeal it and a court finding it invalid. The only way the issue ever will arise is if a party does not appeal against an objectively invalid order (for whatever reason) and also does not comply with it (also for whatever reason). Then only will the party in whose favour an invalid order was granted seek its enforcement before a court. And only when that happens will the court need to decide whether to enforce it or not – as happened in *Motala*. Whether the Constitutional Court affirms *Motala* will have to wait for an appropriate case.

What *Moodley*, *Magnificent Mile* and *Mwelase* all show, is that the Justices of our Constitutional Court take the issues discussed in this article seriously, are alive to the legitimate extent and outer limits of judicial authority, and are committed to ensuring respect for the rule law, and its requirements of legality and certainty, by all State actors.

256 Ibid at para 38.
Surrogacy, Geneticism and Equality: The Case of AB v Minister of Social Development

DENISE MEYERSON

ABSTRACT: Section 294 of the Children’s Act 38 of 2005 provides that a surrogate motherhood agreement will not be valid unless the child born as the result of the agreement is genetically related to at least one of the commissioning parents. This provision was upheld by a majority of the Constitutional Court in AB v Minister of Social Development. I argue in this article that the majority overlooked some troubling constitutional issues. First, I argue that the provision infringes the equality right in s 9(1) of the Constitution, either because the distinction it draws is not rationally connected to the legitimate goal of protecting the best interests of children, or because it serves an illegitimate goal, that of forcibly imposing a contested bionormative conception of the family on people who reasonably disagree. Secondly, I argue that s 294 unfairly discriminates against commissioning parents who would like to enter into a surrogacy agreement but cannot contribute genetic material, thereby infringing s 9(3) of the Constitution. I argue that the AB majority failed to recognise these flaws in the provision because (i) it misidentified the purpose of s 294 and (ii) was too ready to ascribe the predicament of commissioning parents barred from entering into a surrogacy agreement to medical conditions and personal preferences rather than legal discrimination. Its understanding of the constitutional right to equality is accordingly unattractive.

KEYWORDS: Surrogacy, genetic link, right to know genetic origins, infertility, right to equality, unfair discrimination, bionormativity

AUTHOR: Professor of Law, Macquarie Law School, Macquarie University. Email: denise.meyerson@mq.edu.au

ACKNOWLEDGEMENTS: I am grateful to Therese MacDermott, Paul Taylor and two anonymous referees for very helpful comments. Special thanks are due to Max Taylor for invaluable feedback and discussion.
I INTRODUCTION

Prior to the enactment of the Children’s Act 38 of 2005 there was no South African legislation expressly regulating surrogacy agreements. Chapter 19 of the Children’s Act remedies this situation by making it possible for prospective parents who are permanently and irreversibly unable to give birth to a child (‘pregnancy infertile’) to enter into a valid surrogacy motherhood agreement subject to meeting certain requirements. One of these requirements is contained in s 294, which provides that a surrogate motherhood agreement will not be valid unless there is a genetic link between the child born as the result of the agreement and the commissioning parent or parents. In particular, the child must be genetically related to both of the commissioning parents, or, if this is not possible due to ‘biological, medical or other valid reasons’, to one of the commissioning parents. If the commissioning parent is a sole parent, the child must be genetically related to that parent. Thus s 294 prevents commissioning parents who are unable to contribute at least one gamete to conception (‘conception infertile’) from entering into a surrogacy agreement. I will call this differentiation the ‘no-double-donor requirement’.

AB v Minister of Social Development (‘AB’)

Ibid at para 298.

1 AB & Another v Minister of Social Development [2016] ZACC 43, 2017 (3) SA 570 (CC).

2 AB (note 1 above) at para 294.

3 Ibid at para 301.

4 Ibid at para 304.
Turning next to s 12(2)(a) of the Constitution, which guarantees the right to bodily and psychological integrity, including the right to make decisions concerning reproduction, the majority observed that many women do not enjoy security in and control over their bodies. This led the majority to infer that the focus of s 12(2)(a) is ‘on the individual woman’s own body and not a body of another woman’, and hence that the only reproductive decisions protected by s 12(2)(a) are decisions where the body of the person making the decision is affected. Since any decision to enter into a surrogacy agreement would not affect AB’s body (presumably because she would not be contributing a gamete), her right to make decisions concerning reproduction was not restricted by s 294.

Finally, the majority quickly dismissed AB’s arguments that her s 27(1)(a) right to have access to health care services, including reproductive health care, had been violated and that her s 14 right to privacy had been violated. The majority noted that s 27(1) has to be read alongside s 27(2), which obliges the state to take reasonable legislative and other measures, within its available resources, to achieve the progressive realisation of the rights mentioned in s 27(1). It also referred to the finding of the Constitutional Court in *Soobramoney v Minister of Health, KwaZulu-Natal* that the obligation imposed on the state by s 27 in regard to health care is ‘dependent upon the resources available for such purposes, and … the corresponding rights themselves are limited by reason of the lack of resources’. Regarding privacy, although the majority acknowledged that the right to privacy includes the right to make autonomous decisions in respect of ‘intensely significant aspects of one’s personal life’, it stated without further explanation that AB’s right to privacy was not limited by s 294.

The minority would have invalidated s 294. In a judgment written by Khampepe J (Cameron J, Froneman J and Madlanga J concurring), the minority agreed with the majority that s 294 serves the best interests of children and is therefore rationally related to a legitimate purpose. The minority held, however, that s 294 limits the right to psychological integrity by preventing AB and others like her from making decisions concerning reproduction. The minority also held that s 294 draws two distinctions, both of which are unfairly discriminatory. It called these the ‘first differentiation’ and ‘second differentiation’. The first differentiation involves the less favourable treatment of commissioning parents who are conception infertile (This was the differentiation upheld by the majority.) The second differentiation is less straightforward. The minority was concerned about a difference between the law’s approach to double-donation of gametes under the surrogacy scheme by comparison with the in vitro fertilisation (IVF) scheme. The minority referred here to the fact that Regulation 10(2)(a)(ii) of the Regulations Relating to Artificial Fertilisation of Persons, 2012, made under the National Health Act 61 of 2003, allows parents who are conception infertile but who can give birth to a child (ie, are pregnancy fertile) to have IVF treatment using both male and female donor gametes. By contrast, those who are pregnancy infertile and need to enter into a surrogacy agreement would not be able to do so using donor gametes.

---

7 Ibid at para 313.
8 Ibid at paras 314–315.
9 Ibid at para 320.
10 *Soobramoney v Minister of Health, KwaZulu Natal* [1997] ZACC 17, 1998 (1) SA 765 (CC) at para 11.
12 *AB* (note 1 above) at para 323.
13 Ibid at para 104.
14 Ibid at paras 73–97.
15 Ibid at paras 106–119, 127.
agreement in order to have a child are required to contribute genetic material to conception. The minority saw the discrepancy between the surrogacy scheme (prohibiting double-donation of gametes) and the IVF scheme (allowing double-donation) as amounting to another kind of unfair discrimination against AB, in this case on the ground of being pregnancy infertile.\textsuperscript{16} Having found that s 294 limits the rights to equality and psychological integrity, the minority went on to find that the limitations are not reasonable and justifiable in terms of s 36(1) of the Constitution.\textsuperscript{17}

I will focus in this article on the equality issues. Some of my arguments are directed to showing that s 294 should have failed at the first hurdle: the absence of a rational relationship to a legitimate purpose in terms of s 9(1) of the Constitution. The purpose of the no-double-donor requirement is not perspicuous on its face. Both the majority and the minority took the view that it is intended to serve the best interests of children. In its search for a way in which the requirement might achieve this effect, the majority took the view that it satisfies the psychological need of children to have knowledge of their genetic origins. I will argue that other features of the Children’s Act cast doubt on the supposition that this is what the legislature had in mind. An alternative possibility is that the requirement is supposed to serve the best interests of children by creating a strong bond between children and their custodial parents. Although this might have been what the legislature hoped to achieve, I will argue that the exclusion of parents who cannot contribute a gamete is not rationally connected to this purpose, since there is no evidence that the bond between parents and children is weaker in non-genetically-related families. My final suggestion on this matter is that the purpose of s 294 may not be the benign one of serving the best interests of children. I will argue that if s 294 is rationally connected to any purpose, it is the illegitimate purpose of forcibly imposing a bionormative conception of the family – one that regards biological relationships as inherently superior (see part III).

In a further set of arguments, I will suggest that even if the differentiation effected by s 294 does not breach s 9(1) of the Constitution, it nevertheless breaches s 9(3), because it unfairly discriminates against parents who are conception infertile. I therefore disagree with the majority and agree with the minority on this matter. On the other hand, I do not believe that s 294 discriminates against parents who are pregnancy infertile. I therefore disagree with the minority’s arguments in relation to the ‘second differentiation’ (see part II).

Finally, I will suggest that there is a further reason to reject the majority’s interpretation of the purpose of the no-double-donor requirement. As mentioned above, in explaining how s 294 serves the best interests of children, the majority fixed on the importance to children of having knowledge of their genetic origins. I will argue that it is difficult to reconcile this proposition with the IVF regime. The problem here is that the IVF regime not only allows double-donation of gametes, as explained above, but also guarantees the anonymity of gamete donors. This means that children conceived as a result of double-donor IVF treatment will be deprived of identifying information about both their genetic parents. In my view, this creates a problem for the majority’s interpretation of the purpose of s 294, since it implies that the law is selectively concerned with children’s interests in knowing their genetic origins, and hence that the law is morally incoherent. I will argue that judges should be hesitant to interpret a provision in a way which has this implication. In this regard, I am influenced by Ronald Dworkin’s

\textsuperscript{16} Ibid at paras 120–127.

\textsuperscript{17} Ibid at paras 129–213.
principle of integrity in adjudication, which obliges judges to see the law as morally coherent or coherent in principle, so far as possible (see part III). 18

These arguments lead me to conclude that the majority’s understanding of the right to equality under the Constitution is unattractively cramped. In failing to recognise that s 294’s differential treatment of prospective parents who are conception infertile is unfair discrimination, and in being unwilling to interrogate the legitimacy of the state’s purpose, the majority gave us an ungenerous version of the constitutional principle of equality in the service of a dubious legislative goal.

II DOES THE NO-Double-DONOR REQUIREMENT DISCRiminate UNFAIRLY?

In my opinion, both the minority and the majority found too quickly that s 294 passes the s 9(1) test of being rationally related to a legitimate government purpose. Furthermore, once having found to this effect, the majority then found too quickly that s 294 does not discriminate, thereby avoiding the need to determine whether it discriminates unfairly and cutting the equality inquiry prematurely short. I will discuss the s 9(1) issues in part III. In this part, I will assume for the sake of argument that s 294 does further a legitimate purpose, and focus on the discrimination issues, arguing that not only does s 294 discriminate against commissioning parents who cannot contribute a gamete to conception, but that the discrimination is unfair.

The majority found that s 294 does not differentiate on either a specified ground (ie, one listed in s 9(3) of the Constitution) or a ground analogous to the specified grounds (an ‘unspecified ground’). It appeared to regard it as self-evident that the distinction made by s 294 is not based on a specified ground, saying merely that infertility is not mentioned in s 9(3). 19

The minority did not question the majority’s view. 20 Yet disability is a specified ground and most people regard infertility as a form of disability. For instance, the South African Law Reform Commission describes infertility in women as the fifth highest serious global disability among populations under the age of 60. 21 The fact that infertility is not specifically listed in s 9(3) does not support the view that it is not a specified ground, since no specific disabilities are singled out for mention in s 9(3) – presumably because the framers of the Constitution did not think that the protective effect of s 9(3) should extend only to a limited number of specifically mentioned disabilities. These considerations suggest that s 294 differentiates on the basis of the specified ground of disability, in which case it should have been found to be discriminatory and presumptively unfair in terms of s 9(5) of the Constitution. It would then have become necessary to consider whether (i) the presumption of unfairness can be rebutted (by demonstrating that the differentiation does not impair the dignity of conception infertile people), or (ii) the unfairness can be justified in terms of the limitation clause (assuming that it is possible for unfair discrimination to be reasonable and justifiable). The majority bypassed these complex questions when it found that s 294 does not differentiate on the basis of a specified ground.

Having found that s 294 does not differentiate on a specified ground, it was necessary for the majority to consider whether it differentiates on an unspecified ground, so as to come to a

19 AB (note 1 above) at para 298.
20 Ibid at para 105.
conclusion on the discrimination issue. Although the majority did not use the term ‘unspecified ground’, it said that ‘[t]he differentiation will amount to discrimination if the impugned provision authorises unequal treatment of people based on certain attributes and characteristics attaching to them’, and it cited as support for this proposition the established test for differentiation on an unspecified ground as laid down in Harksen NO v Lane (‘Harksen’). According to this test, a differentiation will be based on an unspecified ground if it is based on attributes or characteristics attaching to individuals which have the potential to impair their fundamental dignity or affect them adversely in a comparably serious manner.

In considering whether s 294 ‘authorises unequal treatment of people based on certain attributes and characteristics attaching to them’, the majority made two points, both of which were intended to show that AB’s predicament was not caused by s 294, but rather by factors peculiar to her – her medical condition (which was not a disability) and her choices. First, although the majority conceded that infertile people experience social isolation and marginalisation, it stated that what disqualifies infertile people from entering into a surrogacy arrangement is their biological or medical condition, not the law. Secondly, the majority stated that the inability of people like AB to enter into surrogacy arrangements is due to their personal preferences and hence not a matter of differential treatment based on their ‘attributes or characteristics’. In the majority’s view, AB had the option of finding a fertile permanent partner, whose gamete could then be used for the conception of a child. As the majority judgment put it: ‘[i]f the infertile commissioning parents, or parent, decide not to use the available legal options, they have to live with the choices they make’. Thus it was AB’s ‘personal choice and not her attributes of being infertile or the challenged provision that place[d] her outside of the ambit of s 294’. The majority concluded that the differentiation effected by s 294 does not amount to discrimination. There was therefore no need to consider whether it is unfair.

There are two problems with these arguments. First, although it is true that conception infertile parents suffer from a medical condition which is, as the majority said, not ‘created or compounded’ by the law, it does not follow that it is not the law which disqualifies them from entering into surrogacy agreements. Consider a law which prevents people who suffer from hypertension from becoming airline pilots. Can it be said that because their medical condition is not created by the law, it is not the law which prevents them from becoming airline pilots? This is implausible. Although the barrier is reasonable, it is still a legal barrier. It is likewise the law, not their medical condition, which prevents conception infertile parents from having a child via surrogacy. The minority made the same point in connection with its finding that s 294 limits the right to psychological integrity.

22 AB (note 1 above) at para 298.
23 Ibid at footnote 286.
24 Harksen v Lane NO & Others [1997] ZACC 12, 1998 (1) SA 300 (CC) at para 46 (‘Harksen’).
25 AB (note 1 above) at para 299.
26 Ibid at para 301.
27 Ibid at para 302.
28 Ibid at para 303.
29 Ibid at para 304.
30 Ibid at para 301.
31 Ibid at paras 73–97.
barrier blocking access to surrogacy for conception infertile parents and that it ‘thus becomes the cause of continuing psychological trauma’. 32

The second reason the majority gave for locating the source of the ‘problem’ not in the law, but in people like AB is equally unsatisfactory. This is the proposition that the inability of AB to enter into a surrogacy arrangement was due to her personal choice not to be in a relationship with a (fertile) partner, and hence not a matter of her ‘attributes or characteristics’ at all. The majority’s approach here is reminiscent of certain problematic statements made in Volks NO v Robinson,33 in which the majority of the Constitutional Court upheld a provision which prevented the survivor of a permanent heterosexual partnership from claiming reasonable maintenance from her deceased partner’s estate. In a separate concurring judgment, Ngcobo J emphasised the fact that cohabiting life partners have the legal right to marry, concluding that their inability to access the protective regime which applies to married people is due to their own choice not to marry, not to legal discrimination. 34 It is, however, difficult to accept Ngcobo J’s view that the disadvantages suffered by cohabiting life partners are a simple matter of choice. On the contrary, one partner to an informal relationship (typically the woman) may wish to marry, but the other partner may refuse to do so, frequently so as to evade the legal obligations of marriage.35

The judges in AB similarly exaggerated the extent to which people’s relationships are a matter of personal choice. This is true even of the minority, which accepted the majority’s view that AB had made a deliberate choice not to enter into a relationship with a fertile partner, confining itself to the observation that AB’s decision not to enter into such a relationship was a choice she was entitled to make, since ‘the decision to be single, or follow any other relationship path is deeply personal’. 36 In my view, this concedes too much to the majority. To enter into a surrogacy arrangement, AB would have needed to find an intimate partner with whom she wanted to share her life; her feelings would have had to be reciprocated; and the partner would have had to be both fertile and willing to enter into a surrogacy arrangement. It is difficult to see how these variables can be said to have been under AB’s control, such that her failure to be in a relationship with a person with all these characteristics can be ascribed to her personal preferences. The implausibility of this view is even more obvious if we consider a case where both parties to a marriage cannot contribute a gamete to conception and are therefore disqualified by s 294 from entering into a valid surrogacy agreement. Of course, they could enter into separate surrogacy arrangements if they were to divorce and find fertile partners, but would anyone take the view that if they refuse to avail themselves of this option, they have only themselves to blame? Even radically neoliberal views, which are willing to ascribe many forms of disadvantage to people’s choices, would not be so hyper-voluntarist.

I have explained why I disagree with the majority’s finding that s 294 does not discriminate on the ground of conception infertility. The final question is whether s 294 discriminates unfairly. In my view, the minority was correct to find that it does. Goldstone J explained the concept of unfair discrimination in Harksen, saying that the point of the equality guarantee is to prevent unequal treatment based on criteria which have been used to ‘categorize,
marginalise and often oppress’. He also stated that ‘[i]n the final analysis it is the impact of the discrimination on the complainant that is the determining factor regarding the unfairness of the discrimination’. Infertility is a good candidate for a criterion which meets these tests. As the minority observed, infertility is an emotionally distressing condition and it is also a condition which has been used to stereotype, shame and marginalise people. It is, moreover, obvious that excluding conception infertile people from access to surrogacy services has a highly detrimental impact on them. Section 294 therefore impairs the dignity of conception infertile people and is unfair to them.

Someone might attempt to challenge this conclusion by arguing that people in the situation of AB can avoid the detrimental impact of the no-double-donor requirement by adopting a child. Adoption – it might be said – is the equivalent of surrogacy for them, since they are biologically unable to have a child genetically related to them. This was, in fact, the view of the Ad hoc Parliamentary Committee on Surrogate Motherhood when it recommended the no-double-donor requirement. The Report of the Committee stated:

In instances where both the male and the female gametes used in the creation of the embryo are donor gametes, it would result in a situation similar to adoption, as the child or children would not be genetically linked to the commissioning parent or parents. This would obviate the need for surrogacy as the couple could adopt a child.

The Committee then went on to state that ‘[t]his type of surrogacy was not preferred by most commentators.’ No reasons are given for this ‘preference’, suggesting a moral aversion as such to the use of surrogacy to create children with no genetic tie to their custodial parents. I will return to this possibility later, because in my view the real purpose of s 294 may be to impose a contested moral view about the inherent superiority of biological families. For the moment, it is only necessary to point out that the comparison between adoption and double-donor surrogacy is flawed.

First, as the minority in AB observed, there are psychologically salient differences between double-donor surrogacy and adoption. Double-donor surrogacy allows the commissioning parents to be involved in selecting both the gametes and the surrogate mother, and it also gives them the opportunity to participate in the surrogate mother’s pregnancy. These features of double-donor surrogacy encourage the development of an emotionally significant bond with the child. Secondly, the number of babies available for adoption is limited and the commissioning parents may not even qualify under the adoption rules. The South African Law Reform Commission made these points in a recent Issue Paper which considered whether children should have a legal right to know their biological origins. Contrary, then, to the view of the Ad Hoc Parliamentary Committee, conception infertile parents are not as well served by adoption as by double-donor surrogacy.

37 Harksen (note 24 above) at para 49,
38 Ibid at para 50.
39 AB (note 1 above) at paras 83–89, 106, 112, 120–127.
40 Ibid at para 127.
42 Ibid.
43 AB (note 1 above) at paras 177–185.
44 SA Law Reform Commission (note 21 above) at para 3.32.
It remains only to consider the minority’s view that s 294 unfairly discriminates not only against the conception infertile (the ‘first differentiation’) but also against the pregnancy infertile (the ‘second differentiation’).\(^{45}\) In this regard, the minority relied on a discrepancy between the surrogacy regime and the IVF regime. As mentioned in part I, the Regulations Relating to Artificial Fertilisation of Persons allow parents who are conception infertile but who can give birth to a child (ie, are pregnancy fertile) to have IVF treatment using both male and female donor gametes. As a result, these parents are able to have children who are not genetically related to them. By contrast, pregnancy infertile parents, who need to enter into a surrogacy agreement in order to have a child, are required to be genetically related to the child. The minority thought that the law’s less favourable treatment of conception infertile parents who are also pregnancy infertile is harmful to their dignity, stating that ‘[p]otential users of surrogacy must live with the indignity of knowing that the law gives extra entitlements to those who are not pregnancy infertile, and therefore can use IVF, for no discernible reason supported by argument or evidence’\(^{46}\). The minority therefore concluded that s 294 unfairly discriminates against those who are pregnancy infertile.\(^{47}\)

It is not clear, however, that this discrepancy between the two sets of laws can be used to impugn s 294 as an infringement of the right to equality. The first differentiation is clearly amenable to challenge on this basis because the disqualification is imposed by s 294. Section 294 draws a distinction within the class of pregnancy infertile commissioning parents: those who cannot contribute a gamete are treated less favourably than those who can. Because their less favourable treatment is mandated by s 294, s 294 can be challenged for unfairly discriminating against them. By contrast, the second differentiation is a differentiation within the class of commissioning parents who cannot contribute a gamete: those who can carry a pregnancy are treated more favourably than those who cannot. The former are permitted to have children who are genetic strangers to them, while the latter are not. It is not, however, s 294 which confers the benefit on the former. It is a different law. Since s 294 does not treat commissioning parents who are pregnancy infertile less favourably than those who are pregnancy fertile, it is difficult to see how it can be challenged for drawing an unfair distinction between the two groups of parents.

Is it perhaps the case that s 294 indirectly discriminates on the ground of pregnancy infertility? Does it have a disproportionately adverse effect on prospective parents who are pregnancy infertile? In the case of City Council of Pretoria v Walker, the imposition of a burden on a ground which appeared to be race-neutral (geographical place of residence) was found to be indirectly discriminatory on the grounds of race, because the majority of those who lived in areas where the burden was imposed were white, while the majority of those who lived in areas where the burden was not imposed were black.\(^{48}\) Section 294 does not impose a burden, but it does withhold an opportunity, and withholding an opportunity can also be indirectly discriminatory. This will be the case when the opportunity is withheld on a ground which is facially non-discriminatory but discriminatory in its effect.

The question, then, is whether the requirement that commissioning parents should contribute at least one gamete to conception – a requirement seemingly unrelated to the ability

---

\(^{45}\) AB (note 1 above) at paras 98–101.

\(^{46}\) Ibid at para 123.

\(^{47}\) Ibid at para 127.

\(^{48}\) City Council of Pretoria v Walker [1998] ZACC 1, 1998 (2) SA 363 (CC) at paras 31–33.
to give birth to child – has a differential effect on parents who cannot give birth to a child, amounting to indirect discrimination against them. Assuming that pregnancy infertility is a protected attribute under s 9(3), this question would receive an affirmative answer if pregnancy infertile parents are less likely to be able to contribute a gamete to conception than parents who are pregnancy fertile. This is not, however, the case. By contrast with race and geographical place of residence, which are very closely connected in South Africa, there is no overlap between conception infertility and pregnancy infertility: parents who are pregnancy fertile are just as likely to be incapable of satisfying the s 294 requirement as parents who are pregnancy infertile. Since the provision does not have a disproportionately adverse impact on parents who are pregnancy infertile, it does not indirectly discriminate against them. If anything, the Children’s Act arguably discriminates against parents who are pregnancy fertile, since they are separately disqualified from entering into a surrogacy agreement in terms of s 295 (a) of the Act.

III  IS THE NO-DOUBLE-DONOR REQUIREMENT RATIONALLY CONNECTED TO A LEGITIMATE PURPOSE?

The AB majority took the view that the failure of people like AB to have a fertile partner is a matter of personal preference or choice. Even if this were correct – something that I have disputed – it would still be necessary to explain why conception infertile people like AB should have to give up their single status and find a willing fertile partner in order to enter into a surrogacy agreement. The same question arises where both parties to a relationship suffer from conception infertility. Why should they have to dissolve their relationship and find a fertile partner? People who do not suffer from conception infertility are not required to find or change partners in order to have a child via surrogacy. Suppose, by way of analogy, that the legislature were to decide to restrict access to surrogacy to Christian infertile parents. Religion is a matter of choice, and non-Christians are able to convert to Christianity, but the fact that they can do so would not make this requirement any less arbitrary or irrational. Why, then, is the requirement imposed by s 294 not similarly arbitrary and therefore a breach of s 9(1) of the Constitution?

In their consideration of the s 9(1) issues, both the majority and the minority found that the differentiation effected by s 294 is not arbitrary, saying that the differential treatment for which it provides is a rational way of serving the best interests of children. In coming to this conclusion, the minority omitted to discuss the first differentiation and provided only the briefest of remarks in relation to the second differentiation – the different treatment (according to it) of the pregnancy fertile and the pregnancy infertile. The minority said:

[T]he purpose of s 294 is principally to ensure that the best interests of the child to be born are safeguarded. In the case of surrogacy arrangements, the commissioning parent does not carry the child. This is significantly different to cases where conception is realised through IVF treatment of one of the commissioning parents. That is reason enough to conclude that the differentiation is not arbitrary or capricious. Section 294 is therefore not constitutionally invalid on the basis of section 9(1).49

The majority also had surprisingly little to say in explaining how s 294 is supposed to serve the best interests of the child. It appears that it was influenced by the submission of the Centre for Child Law, which was admitted as amicus curiae. The Centre inferred from the fact that

49  AB (note 1 above) at para 104.
s 294 has the heading ‘Genetic origin of child’ that it was intended to protect the child’s ‘right to know its genetic origin’. Although the Centre conceded that ‘South African laws have not yet formalised the realisation of this right’, it said that international law supports the view that donor-conceived children have a right to know their genetic parents, and that s 294 provides some protection in this regard by ensuring that children born of surrogacy will be able to ascertain the identity of at least one of their genetic parents. According to the Centre, the rationale for recognising the child’s right to know its genetic parentage is that knowing one’s genetic origins is essential to human well-being.

As to how knowledge of one’s genetic ancestry contributes to well-being, the argument standardly made is that children who do not know their genetic parents suffer from identity problems. John Triseliotis, for instance, writes: ‘[i]t can now be claimed with some confidence from the available evidence that there is a psychosocial need in all people, manifest principally among those who grow up away from their original families, to know about their background, their genealogy, and their personal history if they are to grow up feeling complete and whole’. David Velleman makes the stronger argument that actual acquaintance or contact with one’s biological parents is important both for gaining self-knowledge and for identity formation. In relation to self-knowledge, he writes: ‘for information about what I am like as a person, [my biological parents and siblings] are the closest thing to a mirror that I can find’. As far as identity formation is concerned, he claims that one’s biological relations provide the resources with which to construct a meaningful narrative around the events of one’s life, so as to develop a healthy sense of identity. In short, ‘[i]n coming to know and define themselves, most people rely on their acquaintance with people who are like them by virtue of being their biological relatives’. Velleman adds that ‘[n]ot knowing any biological relatives must be like wandering in a world without reflective surfaces, permanently self-blind.’ The phrase, ‘genealogical bewilderment’, used by the British psychologist HJ Sants, is also frequently invoked in this context. Sants used the term to describe (as he saw it) the psychological disorder or confusion about identity experienced by children who do not know their biological origins. I will call this view ‘the Identity Formation Argument’.

It is worth emphasising that the defenders of the Identity Formation Argument believe that ignorance of just one genetic parent is sufficient to give rise to identity problems, since they insist that identity problems arise not only in the adoption context, but also in the context of donor-assisted conception, which generally involves the creation of a child genetically related to one of the parents (e.g., the man is infertile and the woman is artificially inseminated using

---

50 Ibid at para 205.
51 Ibid at para 31.
52 Ibid at para 32.
53 Ibid at para 30.
56 Ibid at 375–376.
57 Ibid at 365.
58 Ibid at 368.
donor sperm). Thus Velleman maintains that children who do not know one of their genetic parents are cut off from one half of an understanding of what they are like.\(^{60}\) He adds:

Their estrangement even from one parent, or from half-brothers and -sisters, must still be a deprivation, because it estranges them from people who would be familiar without any prior acquaintance, people with whom they would enjoy that natural familiarity which would be so revealing about themselves. How odd it must be to go through life never knowing whether a sense of having met a man before is due to his being one's father. How tantalizing to know that there is someone who could instantly show one a living rendition of deeply ingrained aspects of oneself. How frustrating to know that one will never meet him\(^{61}\)

Velleman concludes that it is always wrong to conceive a child using donor gametes, even if only one donor gamete is used and the other custodial parent is genetically related to the child. Sants likewise emphasised that a genealogically bewildered child can be found in any family where one of the natural parents is unknown.\(^{62}\) In short, the defenders of the Identity Formation Argument believe that children need to have a complete picture of their ancestry in order to develop a healthy identity and a true sense of who they are.

Returning now to the reasons why the majority held that the differentiation effected by s 294 is not arbitrary, the majority found that s 294 ‘ensure[s] that the child becomes aware of its genetic origin’.\(^{63}\) It also accepted the Identity Formation Argument, stating that ‘clarity of origin’ is important to children’s self-identity and self-respect, and therefore in their best interests.\(^{64}\) It was consequently beyond question for the majority that s 294 is rationally connected to a legitimate purpose.\(^{65}\) As to whether empirical data support the claim that the absence of a genetic link between commissioning parents and child is harmful to the child, the judges took the view that this was not relevant to the constitutionality of the provision.\(^{66}\) The majority also stated that the rationality of s 294 is as self-evident as the rationality of a provision which disqualifies people who have defective vision or uncontrolled epilepsy from driving.\(^{67}\)

Although the majority readily accepted the Identity Formation Argument, it is by no means uncontroversial. Some writers have pointed to the anecdotal, limited and inconclusive nature of the evidence, as well as the considerable variations in people’s experiences which the evidence reveals.\(^{68}\) Others have argued that any identity problems that children do experience as a result of not knowing their origins are socially induced by virtue of our culture’s over-valuation of biological ties. Thus Sally Haslanger argues that the value our culture places on ‘blood ties’ and ‘naturally’ formed or ‘normal’ families is a socially constructed idea. She describes this

\(^{60}\) Velleman (note 55 above) at 368.

\(^{61}\) Ibid at 368–369.

\(^{62}\) Sants (note 59 above) at 133.

\(^{63}\) AB (note 1 above) at para 254.

\(^{64}\) Ibid at para 290.

\(^{65}\) Ibid at para 293.

\(^{66}\) Ibid at para 299.

\(^{67}\) Ibid at para 291.

idea as the ‘natural nuclear family cultural schema’. Haslanger concedes that this schema can play an important role in forming healthy identities in a cultural context like ours in which it is dominant. She says that children who lack knowledge of their biological parents are ‘left without answers to questions that matter culturally, and this is stigmatizing’. Ensuring that children have knowledge of their origins can consequently help them to overcome the stigma of not ‘fitting’ the schema. This does nothing, however, to challenge the schema, and Haslanger argues that our main goal should be to undermine the dominant ideology of the family, not help people to match it. She gives the analogy of children of inter-racial partnerships who may find it difficult to tell a life story that fits the dominant schema of the ‘Black-White racial binary’. Their difficulties do not mean that inter-racial partners should be prevented from having children. Instead, it means that the schema should be dismantled.

Although these issues are very interesting, there is no space to pursue them here. Instead, I will simply assume for the sake of argument that the Identity Formation Argument is sound and that children need knowledge of their genetic origins in order to develop a healthy identity. I will argue, however, that there are two reasons to reject the majority’s view that the legislature was motivated by these considerations when it enacted s 294. First, I will suggest that when s 294 is read within the context of the Children’s Act as a whole, doubt is cast on the idea that the legislature had children’s identity needs in mind when it required commissioning parents to contribute at least one gamete to conception. Secondly, I will argue that the majority’s supposition about the purpose of s 294 is difficult to reconcile with the IVF regulations, where the law has shown no interest in ensuring that children born as a result of IVF have knowledge of their genetic origins. I will then turn to consider an alternative explanation of why the legislature might have thought that s 294 is a means of advancing the interests of children. The legislature might have thought that when children born of surrogacy are genetically related to at least one of their custodial parents, the bond between children and parents is stronger, and hence that these families function more successfully, whatever the extent of the children’s knowledge on the matter of their origins. This explanation of why the legislature enacted s 294 is more consistent with other provisions in the Children’s Act and is also more consistent with the IVF laws. I will argue, however, that the empirical evidence does not support the view that families function more successfully when there is a genetic relationship between parent and child. The no-double-donor requirement is therefore not a rational way of advancing this objective. Finally, I will suggest that if ensuring genetic connectedness is rationally connected to any goal, it is the goal of imposing a bionormative conception of the family. This is not a legitimate purpose, however, making the majority’s comparison with preventing sight-impaired people from driving inapt.

70 Haslanger (note 69 above) at 179.
71 Ibid.
72 Ibid at 180–181.
73 Ibid at 180.
74 Thaldar also distinguishes between these two views as to the legislative purpose. See Thaldar (note 68 above) at 243.
I will begin by explaining why I do not agree with the majority that the reason why the legislature enacted s 294 was to ensure that children become aware of their genetic origins so as to satisfy their interest in forming a healthy identity. Suppose the commissioning parents have contributed only one gamete – a possibility which is contemplated by s 294 if there are ‘biological, medical or other valid reasons’. Two other features of the legal landscape come into play at this point. First, the law does not impose a duty on the commissioning parents or anyone else to inform the child about the fact that a donor gamete was used in her conception. There is no provision, for instance, requiring this to be recorded on the child’s birth certificate. Secondly, even if the commissioning parents do inform the child that a donor gamete was used, s 41(2) of the Children’s Act prevents the child from discovering the identity of the donor.

In light of this, two scenarios can arise, in neither of which will the child have the knowledge of her origins required to satisfy her hypothesised identity needs. In the first scenario, the child knows that she does not know one of her genetic parents. In the second scenario she falsely believes that she is related to both of her genetic parents. The first scenario will arise if the commissioning parents disclose to the child that she was born as a result of surrogacy and tell her which of her custodial parents is her genetic parent. In this situation, she will be aware of her genetic origin on one side of the family, while also being aware that the identity of her other biological parent is unknown to her and will always remain so by virtue of s 41(2). Although this child has partial knowledge of her origins, her awareness that her knowledge is only partial will be an obstacle to forming a healthy identity in terms of the Identity Formation Argument. As explained above, defenders of the Identity Formation Argument believe that a child who knows that she knows nothing about one of her genetic parents will feel that something important is missing in her life.

In the second (and more probable) scenario, in which the commissioning parents keep the surrogacy agreement secret,75 the parents intend the child to make the false assumption that both of her custodial parents are her genetic parents. She will have one false belief and one true belief as a result. Let us suppose that she falsely believes that her custodial mother is her genetic mother and truly believes that her custodial father is her genetic father. It might be thought that her true belief gives her partial knowledge of her origins, putting her in the same position as the child in the first scenario from the identity formation perspective. This is not, however, correct. The child’s true belief is not justified, since the process by which it was formed was unreliable, and knowledge requires justified true belief, according to the accepted philosophical analysis of knowledge. Suppose someone invents a machine which is supposed to tell you whether your custodial parent is your genetic parent. No-one knows, however, that the machine is unreliable and gives the right answer only half the time. I consult the machine as to whether my custodial father is my genetic father and it tells me that he is. On this occasion, the machine is correct, and the belief I have acquired about my ancestry on my father’s side happens to be true. But do I know this? Clearly, I do not because the machine is unreliable. The same is true if a child is led to believe that both her custodial parents are her genetic parents when this is true only of her father. By contrast with the child in the first scenario, who at least

---

75 Levels of parental disclosure of the use of donor gametes have traditionally been low. Ravitsky cites various studies which show that an ‘overwhelming majority’ of parents do not tell their children that donor sperm or eggs were used in their conception. See V Ravitsky ‘Knowing Where You Come From: The Rights of Donor-Conceived Individuals and the Meaning of Genetic Relatedness’ (2010) 11 Minnesota Journal of Law, Science and Technology 655, 682.
has some knowledge of her origins, this child will lack all knowledge of her origins, because her custodial parents are an unreliable source of information on the matter. She is therefore even less likely than the child in the first scenario to grow up feeling ‘complete and whole’ if the Identity Formation Argument is sound.

Someone might respond to this claim by saying that the ‘knowledge’ of one’s origins required for a healthy identity in terms of the Identity Formation Argument does not have to be fully fledged knowledge in the philosophical sense. True belief, even if unjustified, will suffice. This being the case, the child’s true belief about her father in the second scenario is able to serve her identity needs to the same extent as the child’s partial knowledge in the first scenario. It is difficult to know how to assess the plausibility of this claim, since the defenders of the Identity Formation Argument do not explain what they mean by ‘knowledge’, but it is not necessary to pursue this here, because there is an additional respect in which the child in the second scenario is worse off than the child in the first scenario if the Identity Formation Argument is sound. The child in the first scenario is not in the grip of a false belief. By contrast, the child in the second scenario falsely believes that her custodial mother is her genetic mother. If it is true that ‘knowing who you are requires knowing how you came to be’,76 the self-understanding of the child in the second scenario will be contaminated by her parents’ deception. Although the child in the first scenario will have a gap in her self-understanding, at least she will have a clear-eyed view of what is missing. By contrast, the child in the second scenario will have a false sense of who she is. She will forge a false identity and enjoy an unwarranted sense of self-knowledge built around the wrong genetic narrative. Since she has been misled as to how she ‘came to be’, her confidence that she understands herself will be an illusion based on parental deception.77

In light of the arguments I have made, it is difficult to see how the purpose of s 294 can be to serve children’s interests in constructing their self-identity around knowledge of their genetic parents. If the legislature had really sought to protect this putative interest, it would not have allowed commissioning parents to contribute only one gamete to conception, while simultaneously preventing the child from discovering the identity of the donor of the other gamete, and even permitting the existence of the donor to be kept secret, thereby making it possible for children born of surrogacy to be deceived about their genetic origins and to construct a spurious sense of identity built on an illusion.

To be clear, my argument has not been that s 294 is a poor instrument for achieving the legislative purpose of serving children’s identity needs, let alone that this makes it irrational. An argument to this effect could justifiably be criticised on the ground that rationality analysis in terms of s 9(1) of the Constitution is a limited inquiry, and that courts will not invalidate a scheme merely because it achieves a legislative purpose imperfectly and ‘could have been

76 Ibid at 675.
77 If the Identity Formation Argument is sound, the problem of a ‘false identity’ should also afflict naturally conceived children whose custodial father is not their genetic father and who are not aware of this. Defenders of the Identity Formation Argument should therefore advocate for routine paternity testing to be performed at birth on all naturally conceived children, with the results to be recorded on birth certificates. As Ravelingien and Pennings say, this would put pressure on mothers to inform children about the identity of their real fathers, thereby assuming children ‘genetic truth’ – having a true picture of who they are and where they come from. See A Ravelingien & G Pennings ‘The Right to Know Your Genetic Parents: From Open-Identity Gamete Donation to Routine Paternity Testing’ (2013) 13 The American Journal of Bioethics 33, 34–35. This intervention is the logical next step for defenders of the Identity Formation Argument.
improved in one respect or another’.78 I have not, however, made a means/end rationality argument, let alone one that incorrectly identifies rationality with designing a fully effective means to achieve the legislative purpose. Instead, I have argued that the legislature’s purpose in enacting s 294 could not have been to serve children’s identity needs, because other restrictions and permissions contained in the Children’s Act undermine that construction as a matter of statutory interpretation.79

I turn now to my second reason for thinking that the legislature was not concerned with children’s identity needs when it enacted s 294, namely, the difficulty of reconciling this proposition with the legal regime governing IVF treatment. I will argue that the purpose attributed by the majority to s 294 implies that the IVF and surrogacy regimes embody contradictory moral conceptions – an implication which we should not accept lightly.

I previously discussed the fact that pregnancy fertile parents are permitted to use two donor gametes under the IVF regime, whereas pregnancy infertile parents are obliged to contribute at least one of their own gametes under the surrogacy regime. Although I rejected the minority’s view that the difference in approach under the two regimes means that s 294 is discriminatory, I am now comparing the two regimes from the perspective of the state’s supposed interest in serving children’s psychological needs. When we combine the fact that the IVF regulations allow double-donation of gametes with the fact that donor anonymity is guaranteed by s 41(2) of the Children’s Act, the result is that children born as a result of double-donor IVF treatment will lack knowledge of their genetic origins on both sides. Furthermore, if the IVF treatment is kept secret, these children will falsely believe that they are related to both of their custodial parents when they are related to neither. It is clear, therefore, that in the case of IVF treatment the law is more concerned about the rights of parents and donors to privacy and autonomy than the interest of children in constructing a healthy identity around knowledge of their origins.

Recognising that it needed to explain why the law is concerned about children’s identity needs in the surrogacy context but indifferent to them in the IVF context, the majority offered two (incompatible) responses. First, it suggested that there is a morally relevant difference between IVF treatment and surrogacy. The majority pointed out that IVF treatment is different from surrogacy in that it caters for women who are able to carry a child. Thus, although these women may not be the genetic mother of the child, they do have a ‘gestational link’.80 The majority also took the view that the gestational link is ‘emotionally significant, as it allows the woman to feel that the child is “hers” and to feel that she is a “normal” mother who conceived

78 Jooste v Score Supermarket Trading (Pty) Ltd (Minister of Labour Intervening) [1998] ZACC 18, 1999 (2) SA 1 (CC) at para 16.

79 The minority also thought that the purpose of s 294 is not to give children born of surrogacy knowledge of their origins. However, it came to this conclusion for different reasons. The minority accepted without question the majority’s view that s 294 ensures children acquire knowledge of their origins at least in part, and it also seemed to accept that partial knowledge of origins is sufficient to satisfy children’s identity needs. It held, however, that the partial knowledge ensured by s 294 is a coincidence, not its intended effect. To hold otherwise, according to the minority, amounts to accepting that partial knowledge of origins is consistent with each other. This is because s 41(2) clearly promotes donor anonymity over children’s interests by forbidding donor-conceived children from knowing their origins. Why, then, the minority asked, would the legislature actively seek to protect the interests of a subclass of these children (those born of surrogacy) by ensuring that they know their origins? According to the minority, this would make s 41(2) and s 294 contradictory, rather than complementary: AB (note 1 above) at paras 158–166.

80 Ibid at para 289.
“naturally”). The majority suggested that the feelings created by carrying a child explain why the IVF regulations do not need to insist on a genetic link. In the case of surrogacy, by contrast, a genetic link is desirable in order to compensate for the absence of a gestational link.

However, even if we suppose that the majority is right about the emotional bond created by carrying a child, this still does not explain why it is legitimate for IVF law to deprive IVF children of knowledge of their genetic origins, and even to encourage them to make false assumptions about their origins by allowing their parents to conceal the IVF treatment. If knowing their genetic origins is important to children’s self-identity and self-respect, why would the law lose interest in this vital matter merely because there is an emotional bond between custodial mothers and their non-genetically-related children? How can an emotional bond between mother and child substitute for the psychological importance of knowing one’s genetic origins? The one seems to have nothing to do with the other. Conversely, if an emotional bond can compensate for ignorance of one’s genetic parentage, why did the legislature not recognise this in the context of double-donor surrogacy, where, as explained in part II, the opportunity to select the gametes and the surrogate mother and to participate in the surrogate mother’s pregnancy also create an emotional bond? In short, the fact that IVF treatment caters for women who are able to carry a child does not satisfactorily explain why the law would have sought to give children born of surrogacy knowledge of their origins but denied this knowledge to children born of IVF treatment, since there is no morally relevant difference from the perspective of children’s identity needs between double-donor surrogacy and double-donor IVF.

In its second response, the majority grasped this nettle. Implicitly backtracking on the idea that there is a morally relevant disanalogy between the two forms of assisted reproduction, the majority suggested that IVF law is at fault in unjustifiably neglecting children’s identity needs. Thus, the majority stated:

The risk to children’s self identity and self respect (their dignity and best interests) is, unquestionably, all important. The fact that these rights are placed at similar risk in another context is hardly a reason to find their protection irrelevant.

The majority’s second explanation of the divergence between the two regimes cannot be dismissed out of hand, since it is possible that the law might be selectively concerned with children’s identity needs. Surrogacy law might be intended to serve children’s psychological needs in forming a healthy identity despite the fact that IVF law wrongly repudiates these needs. This would, however, mean that the law is morally incoherent, which is an inference we should be hesitant to draw.

In saying this, I am influenced by Dworkin’s plausible principle of integrity in adjudication, which emphasises the need for judges to see the law as morally coherent or coherent in principle, as far as possible. Dworkin’s principle tells judges to prefer interpretations which enable the law to be understood as expressing a unified moral vision or as speaking with one voice. As he puts it, judges should test their interpretations of particular laws by asking whether they ‘could form part of a coherent theory justifying the network as a whole’. This is not to say that an interpretation of this kind can always be found. Where possible, however, judges should strive

81 Ibid at fn 273.
82 Ibid at para 289.
83 Ibid at para 290.
84 Dworkin (note 18 above) at 245.
to see the total set of laws as morally coherent.\textsuperscript{85} Judges should therefore ‘identify legal rights and duties, so far as possible, on the assumption that they were all created by a single author – the community personified – expressing a coherent conception of justice and fairness’.\textsuperscript{86} The majority’s criticism of the IVF regime for risking the interests protected by the surrogacy regime implies that there is no coherent conception of justice and fairness which underpins both regimes. Yet the majority’s understanding of the purpose of s 294 was not forced on it by unambiguous statutory language. Even if we assume that the generic aim of s 294 is to serve the best interests of children, s 294 might have been intended to serve children’s interests other than by providing them with knowledge of their genetic ancestry. I will shortly explain what this alternative route might be and I will suggest that it is more consistent with the IVF regime’s indifference to children’s identity needs. If these arguments are correct, it follows that the majority’s approach breaches Dworkin’s principle of integrity in adjudication.

In summary, I have argued that there are two reasons to reject the majority’s understanding of the purpose of s 294. First, legal protections for donor anonymity and parental secrecy mean that children raised by parents who qualify under s 294 may not have the knowledge of their origins which is supposed to play such an important role in developing an adequate self-concept. Hence it cannot be the purpose of 294 to provide such knowledge. Secondly, the majority failed to point to any morally relevant difference between the surrogacy and IVF regimes that might help to explain why the IVF regime allows anonymous double-donation of gametes and therefore ignores children’s putative identity needs. Although this is not in itself a decisive consideration, since the law does not necessarily speak with one voice, it does provide a supplementary reason to think that the majority misidentified the purpose of s 294.

What, then, is the purpose of s 294? Assuming for the moment that the no-double-donor requirement is intended to promote good outcomes for children, we need to consider whether s 294 might be intended to achieve this effect in a way which bypasses children’s psychological interests in becoming aware of their origins. Although s 294 does not ensure that children born of surrogacy will have the knowledge they need to construct a healthy identity, it does make it probable that children born of surrogacy will be brought up by at least one genetic parent (subject only to intervening circumstances such as death or divorce). Perhaps the legislature thought that being raised by at least one genetic parent is in the best interests of children – quite apart from what they may know, or not know, or assume, whether truly or falsely, about their genetic origins. As to why this is in the best interests of children, the legislature might have thought that when children are raised by their biological parents, the knowledge that their children are their ‘own’ makes for a ‘natural’ and therefore stronger bond between parents and children, the existence of which promotes more loving and stable families and better outcomes for children.

This possible justification for s 294 is like the ‘clarity of origin’ justification in being based on the best interests of children, but it does not tie the welfare of children to their knowledge of who their genetic parents are. Instead, their welfare is tied to their parents’ knowledge that their children are not genetic strangers to them. At one point the majority seemed to suggest that this might be the justification for s 294, because it said that s 294 serves the purpose of creating a bond between the child and the commissioning parent or parents. The creation of a bond is designed to protect the best interests of the child-to-be-born so that the child has a

\textsuperscript{85} Ibid at 176.

\textsuperscript{86} Ibid at 225.
genetic link with its parent(s).\(^{87}\) Along the same lines, the majority stated that s 294 ‘safeguards the genetic origin of the child’\(^{88}\) (which is different from safeguarding the child’s knowledge of its origin). This justification for s 294 makes more sense of it than the ‘clarity of origin’ justification. Arguably, it also fits better with the law’s toleration of (anonymous) double-donor IVF, if it is true that a gestational link creates an emotional bond between a woman and her non-genetically-related child. Although it is difficult to see how a gestational bond could justify ignoring a child’s identity needs, it might be an acceptable substitute for the hypothesised ‘natural’ bond in genetically-related families.

It is not, however, clear that there is a stronger bond between parents and children in genetically-related families or that these claims about the welfare of children in genetically-related families are true. These are empirical claims and according to surveys of empirical research on the matter they are not well supported. For instance, Lucy Blake, Martin Richards and Susan Golombok comprehensively examine the empirical evidence regarding children’s well-being in different kinds of families and conclude that there is no significant difference in well-being between children in families formed using donor gametes, children who are reared by their genetic parents, and adopted children.\(^{89}\)

Although Blake, Richards and Golombok note that a small number of adopted children experience psychological problems, they say that this is not because they have been reared in a non-biological family. Instead, the slightly poorer outcomes for adopted children are explained by the fact that they are more likely to have experienced adversity in their pre-natal and/or pre-adoption environments. Suffering from neglect, abuse, malnourishment and understimulation in these environments may delay children’s development and interfere with the quality of their attachment relationships. The older they are when they are adopted, the more likely they are to be detrimentally affected by these early stresses.\(^{90}\) At the same time, Blake, Richards and Golombok emphasise that ‘[t]he vast majority of children who are adopted are well within the normal range of well-being and show behavioural patterns that are similar to their non-adopted peers’.\(^{91}\) Blake, Richards and Golombok also discuss studies which compare the well-being of adopted children exposed to pre-adoption adverse experiences with the well-being of children exposed to similarly adverse experiences who are not adopted and who continue to be raised by their biological parents. These studies found that the former do better than the latter, leading Blake, Richards and Golombok to conclude that adoption improves the life chances of children who have been exposed to early stresses.\(^{92}\) This is because most adopted children are moved from an unstable and low socio-economic environment to a nurturing family of higher socio-economic status.\(^{93}\) In so far as families are concerned in which children were conceived using IVF, gamete donation and surrogacy, Blake, Richards and Golombok

\(^{87}\) AB (note 1 above) at para 287.

\(^{88}\) Ibid at para 288.


\(^{90}\) Blake, Richards & Golombok (note 89 above) at 75, 78.

\(^{91}\) Ibid at 76.

\(^{92}\) Ibid.

\(^{93}\) Ibid.
conclude from their analysis of the empirical findings that ‘the families of ARTs [assisted reproductive technologies], including those children who have been conceived with gametes from others, like the majority of adoptive families, function well and their children grow up very much like those in families built through the usual processes of sexual reproduction’.\(^{94}\)

In light of these findings, it is clear that the no-double-donor requirement bears no relationship to the purpose of promoting more loving and stable families, and if this is what s 294 is intended to achieve, it fails to satisfy the rational connection test under s 9(1) of the Constitution. Although, as noted earlier, the legislature does not have to choose the most efficacious means to achieve its goal, its chosen means must make some contribution to the achievement of the goal in order to be rationally connected to it. In the case of s 294, as Donrich Thaldar points out, not only is there no evidence showing that a genetic link is in the best interests of children. There is actually evidence against it.\(^{95}\) It is difficult to understand why the majority took the view that it did not have to consider these issues, but chose to rely instead, as Thaldar observes, on the ‘judges’ personal beliefs regarding the importance of blood ties’.\(^{96}\)

This brings us to the final possibility. It may be that the real purpose of the no-double-donor requirement is not to prevent demonstrable harm to children, but to use the power of the state to advance a particular, bionormative conception of the ideal family form. As Charlotte Witt explains, the bionormative conception takes genetically-related families to be the ‘gold standard’ or ‘Platonic form’ of the family.\(^{97}\) This deeply embedded cultural assumption is also sometimes referred to as the ‘biologic paradigm’. Mianna Lotz observes that the biological paradigm ‘emphasises the primacy, strength and permanence of biologic connection’, in contrast with the ‘second-class status’ of legal kinship, which is viewed as ‘fictive’, ‘fragile’ and ‘impermanent’.\(^{98}\) Empirical research has confirmed the strength and prevalence of the bionormative view that family forms not based on blood ties are less meaningful and less legitimate than those that are.\(^{99}\) Yet defenders of the bionormative conception struggle to pinpoint exactly why genetically-related families are superior.\(^{100}\) As even Velleman concedes, ‘[t]he topic of our biological origins is littered with mythical or symbolic thoughts about blood and bone and seed and such’,\(^{101}\) which do not provide a sound reason to think that biological ties are morally important. Yet Velleman’s own account of the moral importance of biological ties resorts to what he calls ‘universal common sense’\(^{102}\) and the ‘homely truth’\(^{103}\) that ‘[f]irst comes love, then comes marriage, and then the proverbial baby carriage’.\(^{104}\)

\(^{94}\) Ibid at 77. Thaddeus Metz discusses related studies which reach the same conclusion in ‘Questioning South Africa’s “genetic link” requirement for surrogacy’ (2014) 7 South African Journal of Bioethics and Law 34, 35.

\(^{95}\) Thaldar (note 68 above) at 245 (concluding that there is no evidence supporting the existence of a nexus between s 294 and the best interests of the child, and that there is positive evidence that there is no such nexus).

\(^{96}\) Ibid at 253.


\(^{99}\) Ibid at 202.

\(^{100}\) Witt (note 97 above) at 49.

\(^{101}\) Velleman (note 55 above) at 362.


\(^{103}\) Velleman (note 55 above) at 370.

\(^{104}\) Ibid.
and homely truths do not amount to reasoned arguments. Other writers who favour the bionormative conception of the family have similar difficulties in defending it, falling back on ‘what everyone knows about families’, or claiming that our ‘humanness’ is tied to the existence of the biological family.

It is quite likely that bionormative ideas of this kind are the real driving force behind s 294 and that we have arrived here at a purpose which the surrogacy arrangements under the Children’s Act do achieve, namely, the entrenchment of the biologic paradigm and the privileging of some family forms over others. Section 294 can be plausibly explained on the basis that the legislature was operating with an implicit moral hierarchy of the kinds of families which can result from the use of surrogacy. The ranking runs from first-best, to second-best, to altogether unacceptable, with a family’s place in the ranking being based on its proximity to the ‘natural’ family form. This supposition explains the nested structure of s 294. The family closest to the natural form is one in which children born of surrogacy are genetically related to both of their custodial parents. This form is therefore located at the apex of the moral hierarchy, being the outcome which s 294 seeks to secure in the first instance by insisting that, where possible, the gametes of both commissioning parents should be used to effect the conception. Second-best, although still morally acceptable (but only if there are ‘biological, medical or other valid reasons’), is the family in which children born of surrogacy have a genetic connection to one custodial parent. Finally, the family in which neither of the custodial parents is related to the child is seen as an inherently inferior form of family life, sitting at the bottom of the hierarchy, and therefore not to be facilitated. Although it is necessary to tolerate the existence of non-genetically-related families when children who have already been born cannot be cared for by their biological parents, acceptance of the biologic paradigm implies that parents who cannot contribute a gamete to conception should not be allowed to deliberately create ‘second-class’ families by entering into surrogacy agreements.

If the underlying purpose of the law is to prevent the creation of families which the legislature regards as morally inferior, there can be no doubt that s 294 is a suitable means to further it. The question, however, is whether this is a legitimate purpose, as required by s 9(1) of the Constitution. Although the Constitutional Court has on occasion found that legislation has failed this test, the relevant considerations have not always been fully explored. For instance, in two recent cases – Holomisa v Holomisa and Another and Rahube v Rahube and Others – the Court made findings of an impermissible purpose without providing much systematic guidance. The former case dealt with the fact that women married out of community of property under legislation of the former Transkei did not enjoy certain protections on divorce conferred by the Divorce Act 70 of 1979. The Court found it ‘almost impossible to conceive’

105 Witt (note 97 above) at 62.
106 L Kass ‘Making Babies – The New Biology and the “Old” Morality’ (1972) 26 The Public Interest 18, 51. For cogent criticisms of Kass’s views, see Metz (note 94 above) at 36.
107 See Velleman (note 55 above) at 360–361 for the distinction between adopting a child whose ties to its biological parents have ‘been ruptured after conception’ and using gamete donation to intentionally create a child ‘for whom those ties were ruptured antecedently’. Velleman argues that the latter is always morally wrong.
108 Holomisa v Holomisa & Another [2018] ZACC 40 (‘Holomisa’).
109 Rahube v Rahube & Others (‘Rahube’) [2018] ZACC 42 (‘Rahube’).
of a legitimate governmental purpose for the differentiation.\textsuperscript{110} The latter case concerned a provision in the Upgrading Act 112 of 1991, which automatically upgraded land tenure rights under apartheid into full ownership rights. The intention of the Act was to redress the injustices caused by the colonial and apartheid regimes. However, women could not hold land tenure rights under apartheid and were therefore excluded from the upgrading scheme. Since the provision perpetuated an unjust situation created by apartheid legislation, the Court concluded that it lacked a legitimate governmental purpose.\textsuperscript{111}

The most detailed source of guidance in relation to constitutionally impermissible purposes remains \textit{Prinsloo v Van der Linde}, in which Ackermann J stated that in differentiating between people ‘the constitutional state … should not … manifest “naked preferences” that serve no legitimate governmental purpose,’\textsuperscript{112} and added that governmental action should relate to ‘a defensible vision of the public good’.\textsuperscript{113} As noted by Ackermann J, the term ‘naked preferences’ was originally introduced by Cass Sunstein, who used it to capture a theme which unites a number of clauses in the United States Constitution. According to Sunstein, large areas of US constitutional doctrine are directed to ensuring that government action does not result from ‘private pressure’\textsuperscript{114} or a ‘factional take-over’,\textsuperscript{115} but is, instead, the result of a ‘legitimate effort to promote the public good’,\textsuperscript{116} or is justified by reference to ‘some public value’.\textsuperscript{117} Sunstein sought to convey this by saying that these areas of US constitutional law require government action to be grounded in more than naked preferences.

Everyone will agree that governmental action must seek to promote the public good or be justified by reference to a public value in order to count as a legitimate exercise of power in a constitutional state. These terms do not, however, wear their meaning on their sleeve. The \textit{AB} majority readily found that s 294 promotes a public good,\textsuperscript{118} but that was a foregone conclusion in light of its finding that the provision serves the best interests of children – something which is an uncontroversial example of a public good. If, however, the purpose of s 294 is not to serve the best interests of children but to privilege a particular family form on the grounds of its intrinsic moral superiority, the matter becomes more complex. It is impossible to determine whether this is a legitimate purpose without resort to normative theorising directed at giving principled content to the notion of what is ‘public’. In my view, John Rawls provides the most powerful theoretical account of this notion via the distinction he introduces between laws justified by ‘public reasons’ and ‘comprehensive reasons’. I have explained elsewhere why I think that Rawls’s approach is sound.\textsuperscript{119} Here, I will simply summarise his framework and argue that it supports the conclusion that seeking to promote a bionormative conception of the family cannot be justified in terms of public values and therefore amounts to an illegitimate exercise of power.

\textsuperscript{110} Holomisa (note 108 above) at para 24.
\textsuperscript{111} Rahube (note 109 above) at para 43.
\textsuperscript{112} Prinsloo (note 2 above) at para 25.
\textsuperscript{113} Ibid.
\textsuperscript{114} C Sunstein ‘Naked Preferences and the Constitution’ (1984) 84 Columbia Law Review 1689, 1691.
\textsuperscript{115} Ibid at 1690.
\textsuperscript{116} Ibid.
\textsuperscript{117} Ibid at 1692.
\textsuperscript{118} AB (note 1 above) at para 287.
Rawls makes three central claims. First, he maintains that modern democratic societies are inevitably characterised by the fact of ‘reasonable pluralism’. Rawls means by this that reasonable people differ deeply and irreconcilably on many moral, religious, metaphysical and philosophical matters concerning what is valuable in life and gives life meaning. Reasonable people have conflicting ‘comprehensive doctrines’, as he puts it. Secondly, Rawls argues that law-making for comprehensive reasons is an illegitimate exercise of power, or at least that this is the case when the laws concern constitutional essentials (such as the scope of the basic rights and liberties) or matters of basic justice (social and economic equality). According to Rawls, justifying a law by reference to a comprehensive doctrine is equivalent, from the perspective of reasonable citizens who reject the comprehensive doctrine, to having a particular religion forcibly imposed on them. How, then, can the exercise of power be legitimate? How can the state avoid this kind of violation of citizens’ freedom? Rawls’s third claim is that for the exercise of power to be legitimate it must be reasonably justified to everyone. He infers from this that all questions arising in the legislature touching on constitutional essentials and matters of basic justice should be settled by principles and ideals that all citizens can reasonably endorse. Rawls uses the term ‘public’ to capture this idea, saying that arguments regarding these matters should appeal only to public reasons that relate to political values, these being values that are not peculiar to any comprehensive view.

Rawls illustrates his view with several examples. For instance, he discusses possible reasons for infringing people’s religious liberty and says that ‘if we argue that the religious liberty of some citizens is to be denied, we must give them reasons they can not only understand – as Servetus could understand why Calvin wanted to burn him at the stake – but reasons we might reasonably expect that they, as free and equal citizens, might reasonably also accept’. Rawls also discusses different arguments for prohibiting physician-assisted suicide. He observes that some people argue against physician-assisted suicide on the ground that choosing to die is against the will of God. Others argue that if physician-assisted suicide were to be permitted, sick and disabled people might be pressured to end their lives. The former is not an appropriate justification for laws prohibiting assisted suicide, according to Rawls, because it is not a view to which everyone might reasonably agree. Putting this in the terminology used in *Prinsloo*, we can say that prohibiting physician-assisted suicide for religious reasons is more akin to imposing ‘naked preferences’ than seeking to secure a ‘defensible vision of the public good’. By contrast, seeking to protect vulnerable people from being pressured to end their lives is an appropriate justification, not because this justification for prohibiting assisted suicide is one to which everyone will necessarily agree, but because it is a justification to which everyone *might* reasonably agree, since it does not conflict with comprehensive convictions. It

121 Ibid.
123 Rawls (note 120 above) at 137.
124 Ibid at 217.
125 Ibid at 139.
is therefore the right kind of reason to figure in a justification for a coercive law which touches on basic rights and liberties – a public reason.\(^\text{127}\)

The extent to which society should accommodate different familial structures, such as same-sex marriage and polygamy, would seem to be exactly the kind of question which should be decided by reference to public reasons, given the fundamental interests (‘basic rights and liberties’) at stake, such as equality, privacy, intimacy and autonomy. Rawls himself considers justifications for laws regulating the family and assesses them according to whether they are justified by reasons to which everyone might reasonably agree. Thus, he argues that government has no legitimate interest in the particular form of family life, except insofar as it ‘affect[s] the orderly reproduction of society over time’.\(^\text{128}\) What he means by this, as he explains, is that the government’s legitimate interest in the family is confined to securing equal justice for its members by protecting their basic rights and liberties, and freedom and opportunities.\(^\text{129}\) Rawls illustrates the extent of this interest with the example of monogamy, saying that if monogamy were necessary for the equality of women, this would generate a legitimate governmental interest in protecting it. This is because the freedom and equality of women are among the political values of public reason.\(^\text{130}\) By contrast, justifying the protection of monogamy by appealing to its intrinsic value, or its value ‘as such’, would ‘reflect religious or comprehensive moral doctrines’, and would fail to specify a proper interest in the family.\(^\text{131}\) It is perhaps worth emphasising that Rawls is not saying that the state is not entitled to regulate the family. He is merely saying that it is illegitimate to do so for reasons that can only be endorsed by citizens who accept a contested, comprehensive framework.

If the restriction on access to surrogacy agreements contained in s 294 rests on a belief in the intrinsic moral superiority of genetically-related families, or their value ‘as such’, it is clear that the justification would fail Rawls’s test of public reason, because it would not offer reasons which could be shared by all reasonable people regardless of their differences on deeper, comprehensive questions concerning what is of value in life and gives life its meaning. The bionormative conception of the family is an expression of a contested ethical and religious view about what it means to be a good family – a view that stigmatises non-genetic families for reasons which it is not reasonable to think others might reasonably accept. Many reasonable people subscribe to more pluralistic ideas about good families. These opposing reasonable conceptions of the family do not see biological parenthood as inherently superior to social parenthood or blood relationships as the foundation for our humanity. They see value in the social aspects of kinship and the many family forms, both biological and non-biological, that can provide children with love and care and meet their emotional needs. They wish to facilitate, not prevent, the variety of ways – ‘natural’ or not – in which people can become parents.

The minority in AB made related points about the inappropriateness of legally entrenching one form of family life over others and the need to accommodate different family formations in light of the diversity of South African society. However, the minority saw the privileging of genetically-connected families as a matter of discrimination, viewing the genetic link requirement as an affront not only to the dignity of ‘prospective parents, but also of families


\(^{128}\) Rawls (note 126 above) at 779.

\(^{129}\) Ibid at 788–789.

\(^{130}\) Ibid at 793.

\(^{131}\) Ibid at 779.
with adopted children, and our society as a whole’. By contrast, I have suggested that seeking to entrench a particular family form for no reason other than a belief in its intrinsic value amounts to siding with one faction or group in society. The justification in question does not relate to a ‘defensible vision of the public good’. This means that it is a constitutionally impermissible purpose and a breach of s 9(1) of the Constitution.

IV CONCLUSION

I have argued in this article that the no-double-donor requirement contained in the surrogacy regime raises troubling constitutional issues, especially from the perspective of the constitutional right to equality. First, the provision infringes the equality right in s 9(1), either because the distinction it draws is not rationally connected to the legitimate goal of protecting the best interests of children, or because it successfully serves an illegitimate goal, that of forcibly imposing a contested bionormative conception of the family on people who reasonably disagree. Secondly, s 294 unfairly discriminates against commissioning parents who would like to enter into a surrogacy agreement but cannot contribute genetic material, thereby infringing s 9(3). The AB majority failed to recognise these flaws in the provision because (i) it misidentified the purpose of s 294 and (ii) was too ready to ascribe AB’s predicament to medical conditions and personal preferences rather than legal discrimination. Its understanding of the constitutional right to equality is accordingly unattractive.

132 AB (note 1 above) at paras 116–119.
The Constitution as an Instrument of Prejudice: A Critique of *AB v Minister of Social Development*

**DONRICH THALDAR**

ABSTRACT: In *AB v Minister of Social Development*, the applicants challenged extant law that requires surrogacy commissioning parents to use their own gametes for the conception of a child, rather than donor gametes. The majority of the Constitutional Court rejected the challenge on the basis that a child has the right to know his/her genetic origins. The basic premise for this was that knowing one’s genetic origins is important to a child’s development of a positive self-identity. However, the psychological expert opinions before the Court cast doubt on this premise. Why did the majority of the Court disregard the expert evidence and accept the premise? I suggest that the likely reason is that the premise is supported by traditional black South African cultural precepts about kinship formation, which continues to be influential in contemporary South Africa. In this cultural tradition, belonging to a clan is foundational to one’s identity. Belonging, in turn, depends on blood ties with the clan, or, in less rigid varieties of this tradition, depends on at least knowing the child’s genetic origins. Therefore, in this cultural tradition, if a child was conceived using male and female anonymous donor gametes, then belonging to a clan would be impossible, and the child would lack a foundational element of his/her identity. The fact that these traditional cultural precepts about kinship formation continue to be influential in contemporary South Africa cannot support the majority judgment in *AB*, as these cultural precepts constitute prejudice against children based on their social origins. While such prejudice may be beyond the control of the law, the law should never give effect to prejudice. By interpreting constitutional rights such as dignity and the best interests of the child through a lens of prejudice, the Court rendered the Constitution an instrument of prejudice.

**KEYWORDS:** Best interests of the child, dignity, genetic origins, tradition, culture, social origins

**AUTHOR:** Senior Lecturer, School of Law, University of KwaZulu-Natal. Email: thaldard@ukzn.ac.za

**ACKNOWLEDGEMENTS:** I would like to thank Raisa Cachalia and two referees for helpful substantive and editorial assistance on this article. All remaining errors remain mine alone.
I  A LONG TIME AGO, IN A LAND FAR, FAR AWAY…

Shortly before the Interim Constitution entered into force, the case of *Van Rooyen v Van Rooyen*¹ was heard by the then Witwatersrand Local Division of the Supreme Court. The case concerned the revision of the visitation rights of a divorced mother. Subsequent to her divorce from her husband (and father of their children), she entered into a lesbian relationship. The father perceived the children being exposed to the mother’s homosexual relationship as risking psychological harm to the children. The court took specific note of the Bill of Rights that was soon to enter force, and held that it respected the mother’s homosexuality. However, the court also held that should the mother’s lesbian partner share the mother’s bedroom during a visit by the children, it would expose the children to ‘confusing signals’ regarding sexuality that would be ‘detrimental’ to the children.² As such, the court framed the case as a balancing exercise between conflicting rights: on the one hand, the right of the mother not to be discriminated against based on her sexual orientation, and on the other hand, the best interests of the children. The court held that the best interests of the children should be paramount, and consequently that (a) the mother’s lesbian partner may not share the mother’s bedroom during the children’s weekend visits, and (b) the mother’s lesbian partner may not live in the same house as the mother during the children’s holiday visits. However, was this court order really in the best interests of the children?³

Since the time of the *Van Rooyen* judgment, dramatic progress has been made regarding the legal and social equality of homosexual persons. For the contemporary legal analyst, it should be apparent that the *Van Rooyen* court’s reasoning entailed a false dichotomy between the parent’s open homosexual lifestyle, and the best interests of the child. From our current vantage point, it is clear that the *Van Rooyen* court used the best interests of the child as a smokescreen for prejudice.

Our present time may however also be burdened by social and cultural attitudes that may in future – or even presently by more forward-thinking members of society – be regarded as prejudice. What are our contemporary prejudices? Or, stated with more exactness, what are the prevailing social norms that entail inter alia dislike toward an idea or group of persons, but which are not evidence-based? Similar to what happened in the *Van Rooyen* judgment, our courts and lawyers may – consciously or unconsciously – fall prey to prevailing prejudice when interpreting the best interests of the child. This is not only a hypothetical proposition. I suggest that, indeed, it happened in a recent Constitutional Court case, which I introduce next.

II  NOT SO LONG AGO, IN A LAND FAR, FAR AWAY…

The case of *AB v Minister of Social Development*⁴ (hereafter *AB*) centred on how the best interests of the child should be interpreted in a rather unconventional reproductive context, namely surrogate motherhood and the conception of the child by using anonymous donor gametes. AB, the first applicant, had a sad history of failed attempts to become a mother: first by undergoing in vitro fertilisation (IVF) using embryos created from her own eggs and her husband’s sperm; then, given that she was entering menopause and her own eggs were

---

¹ 1994 (2) SA 325 (W) (‘*Van Rooyen*’).
²  Ibid at 328J–330A.
⁴  [2015] ZAGPPHC 580, 2016 (2) SA 27 (GP) (‘*AB HC*’); [2016] ZACC 43, 2017 (3) SA 570 (CC) (‘*AB CC*’).
of insufficient quality, using embryos created from donor eggs and her husband’s sperm; and lastly, following a divorce, using embryos created from donor eggs and donor sperm. Altogether, AB underwent eighteen failed IVF attempts. When IVF proved to be a cul-de-sac, AB started to explore surrogate motherhood as a reproductive avenue. This avenue, she soon discovered, was however legally closed to her: The Children’s Act 38 of 2005, which regulates surrogate motherhood, requires in s 294 that, as a single commissioning parent, she had to use her own eggs for the conception of a surrogacy child. Apart from the fact that it was medically impossible for her to comply with this requirement of the Children’s Act, the bigger normative question was: why is using one’s own genes so important? (Not to mention several legal questions, such as why could AB use donor gametes for self-gestation, but not for surrogate gestation?)

AB challenged the constitutionality of s 294 as being arbitrary, and as infringing six rights in the Bill of Rights: human dignity, equality before the law, non-discrimination, the right to make decisions regarding reproduction, privacy, and access to healthcare. She was joined by the Surrogacy Advisory Group, a non-profit company, as second applicant. The Minister of Social Development, supported by the Centre for Child Law (CCL) as amicus curiae, opposed the application. The best interests of the (future) child took centre stage from the onset: while the applicants argued that the best interests of the child do not require a commissioning parent’s own gametes to be used for the conception of the child, the Minister and the CCL argued the opposite. Importantly, the applicants’ position was informed and supported by extensive evidence provided by child psychology. This included two expert opinions by academics at the University of Cambridge in the United Kingdom who specialise in the psychology of children who were brought into the world using new reproductive technologies, such as donor conception and surrogacy. These experts presented the court with overviews of numerous empirical studies on the psychological well-being of such children. The results consistently showed that a parent–child genetic link is not required for the psychological

---

5 It is interesting to note that women who undergo in vitro fertilisation themselves can elect to use donor gametes for the conception of their children. This choice is not limited to women who cannot, for a medical reason, use their own eggs; this choice is also not limited to just using donor eggs, but can also include donor sperm. Therefore, a woman can choose to be impregnated with an embryo that was created using eggs and sperm from donors of her choice. See AB CC (note 4 above) at para 100. Note, however, that the minority judgment is incorrect in holding that the Regulations Relating to Artificial Fertilisation of Persons of 2012 have been repealed by new Regulations Relating to Artificial Fertilisation of Persons in 2016. The 2016 Regulations were only draft regulations published for public comment, and did not repeal the 2012 Regulations at the time the case was heard.


7 Constitution s 10.

8 Constitution s 9(1).

9 Constitution s 9(3).

10 Constitution s 12(2)(a).

11 Constitution s 14.

12 Constitution s 27(1)(a).

13 AB HC (note 4 above) at para 86.

14 Ibid. See also Golombok expert opinion at para 5 (AB record at 738–739); Jadva expert opinion at para 4 (AB record at 1796).

15 AB HC (note 4 above) at para 86; Golombok expert opinion (AB record at 737–745); Jadva expert opinion (AB record at 1794–1824).
well-being of the child.\textsuperscript{16} The applicants further filed an expert opinion by a local South African clinical psychologist who specialises in the field of new reproductive technologies, and who agreed with the opinions of the Cambridge experts.\textsuperscript{17} In contrast, the Minister filed an opinion by a bioethicist who presented an argument that a parent–child genetic link is essential for the psychological well-being of the child.\textsuperscript{18} This bioethics opinion suffered from numerous demerits, chief among them being that a bioethicist is not qualified to present an expert opinion in the field of psychology.\textsuperscript{19} Eventually, the Minister abandoned her own expert (with good reason) and did not place any reliance on the bioethics opinion.\textsuperscript{20} The CCL did not file any expert opinion. Accordingly, from an evidentiary perspective, the interpretation of the best interests of the child, in the context of the case, should have been clear-cut.

How did the courts decide? Although the Pretoria High Court found in favour of the applicants, a majority of the Constitutional Court found in favour of the Minister and the CCL – disregarding the evidence before it without analysis or good reason. I have previously analysed the \textit{AB} legal saga from an evidentiary perspective, and highlighted this disconcerting snub of the rule of law.\textsuperscript{21} What ideas regarding the best interests of the child were so deeply ingrained in the judicial minds of those deciding the \textit{AB} case, so as to mentally block out the evidence?

In this article, I explore the social normative context of the \textit{AB} decision, identify the social norms that were reinforced by the decision, and consider the constitutional tenability of such social norms and their judicial reinforcement. But first, I familiarise the reader with the interesting argument about the best interests of the child that was the central issue of debate between the parties.

### III THE CHILD’S RIGHT TO KNOW HIS/HER GENETIC ORIGINS

If the evidence is against you, invent a new right. This appears to have been the litigation tactic employed by the Minister in the High Court\textsuperscript{22} and then by the CCL in the Constitutional Court\textsuperscript{23} – and that ultimately secured a majority of the Constitutional Court justices’ votes for its cause. For the purposes of analysis, the right-inventing argument can be formulated as two interconnected syllogisms:

- Knowing one’s genetic origins is important to a child’s development of a positive self-identity (Premise 1a)
- Having a positive self-identity is in the best interest of a child (Premise 1b)
- Therefore: Knowing one’s genetic origins is in the best interest of a child (Conclusion 1/Premise 2a)

\textsuperscript{16} Ibid.
\textsuperscript{17} Rodrigues expert opinion II at para 9 (\textit{AB} record at 2561).
\textsuperscript{18} \textit{AB HC} (note 4 above) at para 85; Van Bogaert expert opinion, part II (revised version)(\textit{AB} record 2577–2595).
\textsuperscript{19} Ibid fn 54; Pretorius expert opinion (\textit{AB} record 2443–2452).
\textsuperscript{20} \textit{AB CC} (note 4 above) at para 202.
\textsuperscript{22} The CCL’s submissions in the High Court focused on comparative law, and did not mention the child’s proposed right to know his/her genetic origins. However, subsequent to the High Court judgment specifically rejecting the proposed right of a child to know his/her genetic origins, the Minister and CCL appear to have each reconsidered their positions. In the Constitutional Court, the Minister abandoned this argument, while the CCL eagerly took up the banner of this argument. See \textit{AB HC} (note 4 above) at para 187.
A CRITIQUE OF AB V MINISTER OF SOCIAL DEVELOPMENT

• The paramountcy of the best interests of a child is a constitutional right (Premise 2b)
• Therefore: A child has the right to know his/her genetic origins (Conclusion 2)

The applicants took issue with the first premise (1a), namely that knowing one’s genetic origins is important to a child’s development of a positive self-identity, and hence rejected the conclusions of the argument.\(^{24}\) The applicants argued that the first premise is a factual statement within the proper province of child psychology, and that there was no psychological expert opinion presented to the court to support this statement.\(^{25}\) In fact, the psychological expert opinions filed by the applicants cast doubt on the first premise. Dr Jadva, one of the Cambridge experts, addressed the question: is not knowing one’s genetic origins likely to impact negatively on one’s psychological well-being (by causing identity problems or otherwise)?\(^{26}\) Dr Jadva started her answer by pointing out that studies show that disclosure of donor-conception elicit a range of positive and negative emotions, depending on the circumstances.\(^{27}\) Finding out that one was donor-conceived earlier in childhood is associated with less negative emotions.\(^{28}\) Dr Jadva offered the following explanation:

> If a child grows up knowing about his/her anonymous donor-conception, it appears that such knowledge is incorporated into his/her identity. In contrast, if a person grows up under the false impression that his/her social parents are also his/her genetic parents, the identity that such person has built during childhood and adolescence is challenged in the event of late disclosure. As such, it is understandable that late disclosure causes more negative feelings than disclosure during childhood.\(^{29}\)

However, the argument in favour of recognising the child’s right to know his/her genetic origins does not focus on a child’s experience of the event of disclosure (of donor-conception), but rather on the broader effect on a child’s self-identity of not knowing his/her genetic origins. Dr Jadva opined as follows in this regard:

> Identity formation happens gradually throughout childhood and especially during adolescence. Multiple factors in a child’s environment impact on this process. Some factors may complicate this process to a greater or lesser degree, but do not necessarily diminish a child’s psychological well-being. Being anonymous-donor-conceived (and hence not knowing one’s genetic origin) seems to be one such factor.\(^{30}\)

The applicants also filed two opinions by Professor Metz, an ethics expert, whose opinions dovetail with Dr Jadva’s opinion.\(^{31}\) Professor Metz argued that:

> Being a ‘coloured’ child in a society that is overwhelmingly either black or white, or being the product of a marriage between a Westernised, Jewish father from an urban metropolis and a sub-Saharan woman from a rural village who believes in African traditional religion, would be ‘ambiguous’ or ‘complicated’ in no qualitatively different a sense than being a child who lacks a genetic relationship with its parents, and yet the state would surely be unjust to prohibit the creation of such children for that reason.\(^{32}\)

\(^{24}\) SAG replying affidavit I at paras 303–309.3 (\(AB\) record at 1648–1650); SAG replying affidavit II at paras 50–59 (\(AB\) record at 2359–2363).
\(^{25}\) Ibid.
\(^{26}\) Jadva expert opinion at para 8 (\(AB\) record at 1799).
\(^{27}\) Ibid at para 17 (\(AB\) record at 1803).
\(^{28}\) Ibid.
\(^{29}\) Ibid at para 30 (\(AB\) record at 1808).
\(^{30}\) Ibid at para 29 (\(AB\) record at 1808).
\(^{31}\) Metz expert opinion I (\(AB\) record at 1950–1967); Metz expert opinion II (\(AB\) record at 2563–2576).
\(^{32}\) Metz expert opinion II at para 14.2 (\(AB\) record at 2571).
Professor Metz pointed out that the logic of the right-inventing argument would entail that the state would be justified in forbidding interracial marital procreation and procreation between people from radically different cultural backgrounds. One can easily expand on this list by thinking of any factor that may complicate a child’s identity formation. Consider the Van Rooyen case: If a child’s mother and father get divorced and the mother enters into an open, intimate homosexual relationship with another woman, is this likely to complicate a young child’s self-identity? And if so, does it constitute a sufficient legal ground to shield the child from being confronted with such a complicating factor? The applicants argued that it does not — it is clear from Dr Jadva’s opinion that psychological complication should not be confused with psychological harm.

Dr Jadva did not confine her opinion to the effect of not knowing one’s genetic origins on self-identity alone (the sole focus of the argument for recognising a child’s right to know his/her genetic origins), but devoted a significant part of her expert opinion to the effect of not knowing one’s genetic origins on children’s psychological well-being in general (including, but not limited to a positive self-identity). With reference to empirical studies, she showed that (a) there is no difference in the psychological well-being of donor-conceived children and naturally conceived children, and that (b) there is no difference in the psychological well-being of donor-conceived children who know the identity of their donors, and donor-conceived children where anonymous donors were used. Accordingly, Dr Jadva answered the question posed to her as follows: Not knowing one’s genetic origins does not appear likely to impact negatively on one’s psychological well-being.

Dr Jadva’s conclusion, which was based on quantitative empirical studies, was confirmed from a clinical perspective by Ms Rodrigues, a South African clinical psychologist who works with surrogacy, infertility, and gamete donor selection. Ms Rodrigues elaborated as follows:

In my observation, donor-conceived children have not manifested any discernable higher incidence of psychological problems than children in the general population. [...] It should be considered that IVF using donor gametes has been practised in South Africa for over a generation, and has become generally accepted medical practice in South Africa. During all this time, I am not aware of any report in the literature or at psychology conferences or workshops that contradicts my observation [...] above.

These expert opinions — one would think — should have convincingly countered the argument for recognising a child’s right to know his/her genetic origins. However, the Constitutional Court majority disregarded this evidence. The proffered reason was that the Court, and not experts, is the ultimate authority on questions regarding the validity of legislation, and that the Court should arrive at its own independent evaluation of the issues before it. What these legal rules do support is not blindly accepting experts’ conclusions, but critically engaging with the evidence, evaluating its credibility and reliability, and then applying it to the legal question

33 Ibid.
34 Jadva expert opinion at paras 20–27 (AB record at 1804–1807).
35 Ibid.
36 Jadva expert opinion at para 27 (AB record at 1807).
37 Rodrigues expert opinion II at para 9 (AB record at 2561).
38 Rodrigues expert opinion I at para 4 (AB record at 853).
39 Rodrigues expert opinion II at paras 8.1–8.2 (AB record at 2560–2561).
40 AB CC (note 4 above) at para 269.
41 Ibid.
before the Court. But clearly being the ultimate authority and engaging in independent evaluation do not support just disregarding expert evidence properly before the Court.

Yet, in an indirect and deeply perverse way, the expert evidence did influence the majority judgment. This is how: the conclusions in the expert opinions filed by the applicants are general statements that point to probability, not certainty. In typical scientific fashion, the Cambridge experts used cautious language and qualified their opinions with words such as ‘likely’, ‘not necessarily’, and ‘may’. Similarly, Ms Rodrigues also carefully avoided absolute statements. Consider for instance the following core statement in the expert opinion by Ms Rodrigues: ‘In my observation, donor-conceived children have not manifested any discernible higher incidence of psychological problems than children in the general population’. This means that donor-conceived children generally do not manifest more psychological problems than children in the general population. It does not mean that donor-conceived children never manifest psychological problems; there is a possibility, however slight, that some donor-conceived children may indeed suffer from psychological problems.

These probabilistic, non-absolute statements formed the basis of the applicants’ argument in court – the applicants’ position was that where a child does not know his/her genetic origin, it is unlikely to impact negatively on a child’s self-identity or psychological well-being. The majority of the Constitutional Court pounced on the non-absolute nature of the applicants’ argument as a weakness in the applicants’ case:

The applicants did not dispute that the clarity of origin may be important to a self-identity and self-respect of the child. […] The risk to children’s self-identity and self-respect (their dignity and best interests) is, unquestionably, all important.

This is a confounding development in our law: instead of requiring the state party to show that there is a rational nexus between the impugned provision and a legitimate government purpose, the fact that the party that challenges the impugned provision cannot with absolute certainty prove that there is no such nexus, is perceived as establishing such a nexus. In more general terms: if you cannot disprove a proposition with absolute certainty, it proves the proposition. Thus, although not explicitly but certainly implicitly, the child’s purported right to know his/her genetic origins was recognised by South Africa’s highest court. But what deeply rooted beliefs could have caused the majority of the Constitutional Court to negate the rules of evidence and logic in such obvious and extreme ways when it came to the issue of a child’s blood ties with his/her parents?

42 See, for instance, Teddy Bear Clinic for Abused Children v Minister of Justice and Constitutional Development [2013] ZACC 35, 2014 (2) SA 168 (CC), where the Constitutional Court analysed the psychological expert opinion at length, and explicitly based its decision on such evidence. An overview of our law regarding expert evidence is presented in Twinie v Naidoo [2017] ZAGPJHC 288, [2018] 1 All SA 297 (GJ) at para 18.
43 I analyse this issue in more detail in Thaldar (note 21 above).
44 Rodrigues expert opinion II at para 8.1 (AB record at 2560).
45 AB CC (note 4 above) at para 290 (emphasis added).
46 Ibid at para 294 (‘[T]he substance below the surface is the need for a genetic link between a child and at least one parent […] clarity regarding the origin of a child is important to the self-identity and self-respect of the child.’)
IV  THE UNARTICULATED DEEPEST PREMISES

In a recent lecture, Cameron J (part of the AB minority) observed as follows:

[T]he duty of constitutional guardianship and exposition cannot and does not stop at securing democratic political processes. More is entailed. For in applying any form of law judges are called on to do more than only determine the outcome of disputes before them. They are required to explain the normative framework that impels the answers they give, and sometimes to expound its deepest premises.47

Perhaps the ‘deepest premises’ of the ‘normative framework’ that impelled the answers given by the majority of the Constitutional Court in AB are reasonable, convincing, and aligned with the values of the Constitution – and the majority of the Constitutional Court just failed to articulate such premises sufficiently? There are several clues pointing to such premises. First, the majority, relying on the Children’s Act, held that ‘keeping the connection with extended family, culture and tradition is indeed part of the factors showing where the best interests of the child lie’.48 This statement in isolation does not assist the current analysis (nor the majority’s position), given that the Children’s Act also provides that a surrogacy child is the child of the commissioning parents,49 not the gamete donors;50 accordingly, a hypothetical surrogacy child who was conceived using male and female anonymous donor gametes will have an extended family, culture and tradition – the extended family, culture and tradition of his/her commissioning parents (and not the extended family, culture and tradition of his/her gamete donors). However, this statement by the majority must be seen in the light of the other clues: during oral argument, Moseneke DCJ (who was part of the majority) remarked that a parent-child genetic link is of vital importance in ‘certain cultures’, and twice referred to a ‘Zulu man’ as a supposed example of someone for whom such a genetic link would carry such importance. In the same vein as Moseneke DCJ’s remarks, the minority judgment refers to ‘the African custom that requires that a boy must know the genetic origin of his father to form a part of the community’.51 From a reading of the majority judgment as a whole, it is clear that what the majority had in mind was, in fact, the child’s connection with the extended family, culture and tradition of his/her genetic parents (contra the provisions of the Children’s Act).

It would appear that the deepest premises underlying the majority judgment are certain cultural norms that emphasise the importance of genetic origins in the conception of what constitutes a child’s family, culture and tradition. These cultural norms deserve analysis. There is a growing body of research on public perceptions in South Africa about the adoption of

48  AB CC (note 4 above) at para 300.
49  Children’s Act s 297(1)(a) (‘The effect of a valid surrogate motherhood agreement is that— (a) any child born of a surrogate mother in accordance with the agreement is for all purposes the child of the commissioning parent or parents from the moment of the birth of the child concerned.’)
50  Children’s Act s 1 (‘[P]arent’, in relation to a child, […] excludes— (b) any person who is biologically related to a child by reason only of being a gamete donor for purposes of artificial fertilisation.’)
51  AB CC (note 4 above) at para 197.
genetically unrelated children\(^{52}\) – often abandoned babies whose genetic origins are unknown.\(^{53}\) The results of this research illuminate the cultural norms regarding kinship formation are prevalent in contemporary South Africa. In the following paragraphs, I present an overview of these cultural norms, based on this research.

In traditional black South African culture, ancestral spirits play an important role. One particular area where belief in ancestral spirits features prominently is in kinship formation. When a child is born, tradition demands that the child must be ritually introduced not only to living clan members, but also to the clan ancestors, in order to secure the ancestors’ acceptance of the child into the clan.\(^{54}\) As a child grows up, the importance of belonging to the clan, which is perceived as including the living clan members and their ancestors, is inculcated in the child’s psyche; belonging to the clan – one’s social and spiritual roots – becomes foundational to a child’s identity. Accordingly, the knowledge that one has been accepted by the clan’s ancestors and that one therefore belongs to the clan are critical factors for a child to develop a positive self-identity.

The ancestors’ acceptance of a child into the clan is not only a matter of performing a ritual, but pertinent requires that the child must have blood ties with the clan; without such blood ties, the ancestors will not accept a child as a member of the clan.\(^{55}\) Upsetting the continuity of blood ties between the ancestors and the next generation is perceived as improper

---


53 Blackie (note 52 above).

54 Gerrand & Nathane-Taulela (note 52 above) at 58 (Most black South African families still practise ceremonies to introduce the biological child to living relatives and ancestral spirits, so that the child can develop a sense of belonging and identity.) Gerrand (note 52 above) at 69 (If rituals like ‘imbeleko’ – introducing the new-born baby to the ancestors – are not performed according to the customs of the paternal clan, misfortune and bad luck are said to follow the child for the rest of his/her life.)

55 Gerrand (note 52 above) at 142. (Observes that ‘in terms of ancestral beliefs, family systems have rigid boundaries based on blood ties.’) Gerrand’s view is echoed by the KwaZulu-Natal Commissioner for Traditional Leadership Disputes and Crimes, Professor Jabulani Mphalala, who has stated that ‘it would take years before there was a flexibility of mind about adoption among most South Africans. [...] Ancestral spirits look after their relatives and no-one else. In our religion, in our culture, this thing is ring-fenced’. C Dardagan ‘Red Tape Slowing down Adoptions’ IOL (21 February 2014), available at https://www.iol.co.za/lifestyle/family/parenting/red-tape-slowing-down-adoptions-1650829.
and offensive\textsuperscript{56} – and can invoke retribution by the ancestors.\textsuperscript{57} At most, an unrelated child can be taken into foster care, subject to preference being given to other children who are blood family.\textsuperscript{58} This belief in the rigid demarcation of the family as a social structure through blood ties is clearly irreconcilable with the adoption of unrelated children and surrogacy using male and female anonymous donor gametes. Such practices would expose the adoptive or commissioning parents – and even the unrelated adoptive or surrogacy child – to the wrath of the ancestors.

Accordingly, the situation of a hypothetical surrogacy child who is conceived using male and female anonymous donor gametes (and who is therefore unrelated to the commissioning parents) would indeed be bleak. Apart from being the embodiment of an offence against the ancestors and an entire cultural tradition, the child could never truly become part of the clan, and given that clan membership is emphasised as a foundational element of a person’s identity, the child irredeemably lacks a foundational element of his/her identity. Moreover, on a spiritual level, such a child can never know his/her genetic ancestors, with unpleasant and even punitive consequences for the future happiness of the child. Clearly, in traditional black South African culture, intentionally wanting to create such a child is evil. Note that this is exactly what AB wished to do, and what the applicants in \textit{AB} intended to legalise.

A less rigid version of this belief entails that a child without blood ties can be accepted by the ancestors \textit{provided} that the child’s ancestry is known, and provided that the ancestors are informed of the child’s ancestry during the ritual introduction of the child to the ancestors.\textsuperscript{59} However, this less rigid version is still problematic from the perspective of adoption of unrelated children whose genetic origins are unknown, such as abandoned babies and surrogacy using male and female anonymous donor gametes. In instances like these, where a child’s genetic origins are unknown, only a very relaxed version of the belief in ancestors would allow for the child to be introduced to the ancestors and accepted in the clan.\textsuperscript{60}

\textsuperscript{56} Mokomane & Rochat (2010) (note 52 above) at ix; Rochat, Mokomane & Mitchell (note 52 above) at 124 (A study participant remarked: ‘When you are born, there are certain things that ancestors require of us. They know who our child is and where he is. Just imagine if you adopt a Biyela child and join the child to the Mthembus. There will be war between the Biyela and Mthembu ancestors, both ancestors will fight over who owns the child.’)

\textsuperscript{57} Blackie (note 52 above) at 6 (Observes, in the context of the adoption of abandoned children, the common belief that an adopted child’s lack of connection with his/her genetic ancestors will result in a troubled and unfulfilled life for the child, and that the adoption would also cause problems with their ancestors for the adoptive parents.) Rochat, Mokomane & Mitchell (note 52 above) at 124 (Observe that ‘children deprived of their roots would lose contact with their ancestors, with unpleasant, punitive consequences for the future happiness of the child’); Gerrand (note 52 above) at 208 (Cautions that unrelated adoption would ‘probably expose the adopted child to risk’; study participants emphasised that in traditional African culture, trying to integrate a ‘foreign’ child into a kin system can invoke retribution by the ancestors.)

\textsuperscript{58} Gerrand (note 52 above) at 140–141.

\textsuperscript{59} Martin & Mbambo (note 52 above) at 43 (Quote traditional leaders who participated in their study as remarking that: ‘It is not possible to adopt a child whose surname you don’t know. How can you slaughter a goat for that child? Whose ancestors are you going to call when you do not know his/her source?’) See Rochat (note 52 above) at 124 (Ritualistic practices to appease ancestors and allow non-kin placement of children require information on the child’s genetic origins.) See also Gerrand (note 52 above) at 131.

\textsuperscript{60} Gerrand (note 52 above) at 133 (Some study participants – all participants were black South Africans – regarded denying adoptable children the opportunity to enter a loving family because of staunch ancestral beliefs as ‘ridiculous.’)
How prevalent is the belief in ancestors among black South Africans in contemporary South Africa? Similar to the way in which certain pre-modern Western cultural elements, most notably Christianity, continue (one could argue anachronistically) to exercise influence on the modern Western world, certain traditional black South African cultural elements also continue to exercise influence among contemporary black South Africans.\footnote{Ibid at 131 (A study participant remarked: ‘Ninety-nine per cent of people in our culture, they believe in ancestors.’)} Although many black South Africans have adopted a modern lifestyle, this does not mean they have abandoned their traditional beliefs. The traditional construct of kinship formation in particular is something that many black South Africans continue to strongly identify with.\footnote{Ibid at 208 (‘[M]any well-educated, black South Africans identify with a traditional African cultural construct of kinship.’)} This is vividly illustrated by the stark difference in adoption rates between South Africa’s white and black populations. A black person is about nine times less likely to be an adoptive parent than a white person\footnote{Mokomane & Rochat (note 52 above) at 353.} – despite the high number of black children waiting to be adopted.\footnote{Blackie (note 52 above) at 8 (In November 2013, the Registry of Adoptable Children and Parents (RACAP), showed that of the unmatched children, 398 were black, 3 were white, and 9 were termed ‘mixed’ race.)}

Given these insights into the deepest premises of the normative framework of contemporary black South Africans, the question must be asked: was the finding by the majority of the Constitutional Court in \textit{AB} not, after all, reasonable and ultimately correct? These insights can be fertile soil for an argument against the applicability of the psychological expert opinions presented by the applicants, and in support of the constitutionality of s 294 (the impugned provision). Such an argument that could have provided support for the position taken by the majority of the Constitutional Court in \textit{AB}, can be systematically constructed as follows (‘the tradition-based argument’):

- In traditional black South African culture (Value A), belonging to a clan is foundational for one’s identity, and (Value B) belonging to a clan depends on blood ties with the clan, or (in less rigid varieties of tradition) depends on knowing the child’s genetic origins. (Premise 1)
- Therefore (because of Premise 1), in traditional black South African culture, if a child was conceived using anonymous male and female donor gametes, (because of Value B) belonging to a clan would be impossible, and (because of Value A) the child will lack a foundational element of his/her identity. (Conclusion 1/Premise 2)
- Traditional black South African cultural precepts about kinship formation are influential in contemporary South African society. (Premise 3)
- Therefore (because of Premise 3), children in contemporary South Africa are likely to grow up in a community where traditional black South African cultural precepts about kinship formation are influential. (Conclusion 2/Premise 4)
- Therefore (because of Premises 1 and 4), children in contemporary South Africa are likely to internalise Values A and B. (Conclusion 3/Premise 5)
- Therefore (because of Premises 2 and 5), children in contemporary South Africa, if conceived using anonymous male and female donor gametes, are likely to feel that they lack a foundational element of their identity. (And since identity is part of one’s dignity, it also follows that such children will also feel that their dignity has been violated). (Conclusion 4/Premise 6)
• Section 294 was designed to avoid a situation where surrogacy children can be conceived using anonymous male and female donor gametes. (Premise 7)
• Therefore (because of Premises 6 and 7), s 294 protects the best interests of the child (including the child’s dignity). (Conclusion 5/Premise 8)
• The paramountcy of the best interests of a child is a constitutional right; dignity is a constitutional right (Premise 9)
• Therefore (because of Premise 8 and 9), s 294 is aligned with the Constitution. (Conclusion 6)

If this argument is accepted, an attack on the constitutionality of s 294 is therefore a balancing exercise between the rights of commissioning parents, like AB, who wish to use male and female anonymous donor gametes, and the best interests of the future child. But there is a special moral and legal relationship between these supposed opposing parties: parents are morally and legally expected to care for their children, and unselfishly act in their children’s best interests. Accordingly, the balancing scale must tilt in favour of the best interests of the future child. The attack on the constitutionality of s 294 must fail.

Does this settle the matter? Was the majority judgment in AB correct to reject the attack on the constitutionality of s 294? In the following sections, I analyse four potential counter-arguments to the tradition-based argument. While there are degrees of similarity between the first three counter-arguments and the arguments contained in the AB minority judgment, the fourth counter-argument – which I suggest is the most profound argument against the tradition-based argument – was not canvassed in any of the AB judgments and presents a new way of thinking about the issues in AB.

V COMMONSENSE PROBABILITIES

Consider the ‘Zulu man’ referred to by Moseneke DCJ as the stereotype of a staunch adherent of traditional black South African culture. It strikes one as counter-intuitive that the ‘Zulu man’ – if he is such a staunch adherent to traditional black South African culture – will ever contemplate using anonymous male and female donor gametes to beget a child. Stated in terms of probability, the probability of a surrogacy child conceived using male and female anonymous donor gametes being born to parents who adhere to traditional black South African cultural precepts about kinship formation is extremely unlikely.65

Also, just as an intended parent who is an adherent of traditional black South African cultural precepts about kinship formation is unlikely to, in the first place, conceive a child using male and female anonymous donor gametes, so an intended parent who elects to conceive a child using male and female anonymous donor gametes is likely to be an adherent of modern precepts about kinship formation that do not emphasise or require blood ties, and are likely to raise the child in a like-minded fashion. In short: AB is not a Zulu man – at least not the stereotypical ‘Zulu man’ intended by Moseneke DCJ.

65 The AB minority observed that parents who are adherents to traditional black South African cultural precepts of kinship formation are not compelled to use male and female anonymous donor gametes to beget a child against their cultural beliefs. See AB CC (note 4 above) at para 197. However, the AB minority does not take this observation further into the context of balance of probabilities – the civil standard of proof. The AB minority’s point of departure is that there should be a case-by-case approach, and that diversity should be respected. As such, the AB minority’s position corresponds most closely with the counter-arguments analysed under the headings ‘VI The Case-by-Case Approach’ and ‘VII The Value of Diversity’ below.
Now consider Conclusion 2/Premise 4 of the tradition-based argument. Children in contemporary South Africa are *likely* to grow up in a community where traditional black South African cultural precepts about kinship formation are influential. This may hold true for children in general, but not for the subgroup of children that are relevant for purposes of the tradition-based argument, namely children conceived using anonymous male and female donor gametes. For this relevant subgroup of children, the exact opposite of Conclusion 2/Premise 4 holds true: children conceived using male and female anonymous donor gametes are extremely *unlikely* to grow up in a community where traditional black South African cultural precepts about kinship formation are influential, and are far more *likely* to grow up in a community that values modern, inclusive precepts about kinship formation. Accordingly, the tradition-based argument collapses.

VI THE CASE-BY-CASE APPROACH

One could also argue that even the *slightest possibility* of a child conceived using male and female anonymous donor gametes growing up in a community where traditional black South African cultural precepts about kinship formation are influential is still unacceptable, given the *devastating impact* that it would have on a child’s self-identity and dignity. This calls for a precautionary approach that would entail the following reformulation of the tradition-based argument:

- Premise 1 and Conclusion 1/Premise 2 remain unchanged from the original formulation of the tradition-based argument.
- There is a *possibility*, however slight, that a child conceived using male and female anonymous donor gametes may find himself/herself growing up in a community that adheres to traditional black South African cultural precepts about kinship formation. (Premise 3, reformulated)
- Therefore (because of Premise 1), children who grow up in a community that adheres to traditional black South African cultural precepts about kinship formation are likely to internalise Values A and B. (Conclusion 2/Premise 4, reformulated)
- Therefore (because of Premises 2, 3 and 4), children who grow up in a community that adheres to traditional black South African cultural precepts about kinship formation, if conceived using anonymous male and female donor gametes, are likely to feel that they lack a foundational element of their identity. And, since identity is part of one’s dignity, it also follows that such children will also feel that their dignity has been violated. (Conclusion 3/Premise 5, reformulated)
- Section 294 was designed to avoid a situation where surrogacy children can be conceived using anonymous male and female donor gametes – hence eliminating the possibility contemplated in Premise 4. (Premise 6, reformulated)
- Therefore (because of Premises 5 and 6), s 294 protects the best interests of the child, including the child’s dignity, by eliminating the possibility contemplated in Premise 3, and hence eliminating the negative consequences contemplated in Premise 5. (Conclusion 4/Premise 7, reformulated)
- The paramountcy of the best interests of a child is a constitutional right; dignity is a constitutional right (Premise 8)
- Therefore (because of Premise 7 and 8), s 294 is aligned with the Constitution. (Conclusion 5, unchanged from the final conclusion of the original version of the tradition-based argument)
This reformulated version of the tradition-based argument, based on the precautionary principle, successfully addresses the commonsense probabilities counter-argument. However, consider the following question: can the possibility that some exceptional cases may occur where commissioning parents, who adhere to traditional black South African cultural precepts about kinship formation, counter-intuitively want to have a surrogacy child using male and female anonymous donor gametes, serve as basis for a *blanket ban* on all commissioning parents using male and female anonymous donor gametes? Such a blanket ban would be overly broad and disproportionate to the object that it wants to achieve. If one accepts the precautionary principle in the present context, it can serve as a basis to require commissioning parents in *individual* surrogacy agreement confirmation applications where the intention is that the surrogacy child will be conceived using male and female anonymous donor gametes, *not* to bring up the child in traditional black South African culture or any other culture that places similar emphasis on genetic lineage as being determinative for a person’s identity. This examination of the best interests of a child on an *individual* basis would be aligned with Constitutional Court precedent, and would adhere to the Constitution’s proportionality principle. Conversely, accepting the precautionary principle cannot serve as a basis for a blanket ban, particularly where the rationale for the ban does not exist.

This identifies the logical mistake in the reformulated version of the tradition-based argument based on the precautionary principle, namely that Conclusion 4/Premise 7 is an overly broad conclusion from its premises. The blanket ban on using male and female anonymous donor gametes embodied in s 294 cannot be said to protect the best interests of the child in all or even most cases, but only in some exceptional, counter-intuitive cases. In all but the most exceptional, counter-intuitive cases, such a ban does not contribute to its objective of protecting the best interests of the child, and simply limits the autonomy of commissioning parents. Accordingly, even the reformulated version of the tradition-based argument that is based on the precautionary principle fails to justify s 294.

Note that the examination of the best interests of a child on an individual basis with relation to the impact of being conceived using male and female anonymous donor gametes would not require any new legislation or complex reformulation of s 294. Chapter 19 of the Children’s Act already contains a general provision, s 295(e), which allows for any issue that may impact on the best interests of the child to be considered by the court.

**VII THE VALUE OF DIVERSITY**

The arguments against the tradition-based argument (in both its original and reformulated versions) have thus far focused on *actual* cultural diversity, logic, and the principle of proportionality. These arguments can be amplified by considering diversity as *value*. The preamble of the Constitution states that ‘South Africa belongs to all who live in it, united in

---

66 *AD v DW*[2007] ZACC 27, 2008 (3) SA 183 (CC) at para 55 (Child law is an area that abhors maximalist legal propositions that preclude or diminish the possibilities of looking at and evaluating the specific circumstances of the case.)

67 *AB CC* (note 4 above) at para 193 (What is in a child’s best interests is a flexible inquiry which must be determined on the facts of each particular case.)

68 Ibid at para 192 (Section 295(e) of the Children’s Act mandates the High Court on every occasion it decides whether to confirm an agreement to engage with the value judgement of whether the agreement would be in the best interests of the child to be born.)
our diversity’. A rich body of human rights jurisprudence embraces South Africa’s cultural diversity and supports the idea of a ‘right to be different’. This nostrum suggests that persons should not be forced to subordinate themselves to the cultural and religious norms of others. This orientation is particularly applicable to kinship formation. In *Du Toit v Minister for Welfare and Population Development*, the Constitutional Court held that: ‘[f]amily life as contemplated by the Constitution can be provided in different ways and that legal conceptions of the family and what constitutes family life should change as social practices and traditions change.’ Similarly, in *Minister of Home Affairs v Fourie*, the Constitutional Court held that ‘South Africa has a multitude of family formations that are evolving rapidly as our society develops, so that it is inappropriate to entrench any particular form as the only socially and legally acceptable one.’

The tradition-based argument fails to provide for evolving concepts of family formation as contemplated in *Du Toit* and *Fourie*, but rather does the opposite: it legally entrenches genetic lineage as a condition for surrogacy family formation. On the issue of a genetic lineage (or rather the lack thereof), the Constitutional Court held as follows in *Fourie*: ‘It is demeaning to adoptive parents to suggest that their family is any less a family and any less entitled to respect and concern than a family with procreated children.’ Clearly, a family with genetically unrelated children should be entitled to the same respect and concern as any family with naturally conceived children. Instead of fixating only on cultural contexts where such a novel type of family formation would be frowned upon and where the child may suffer as a result, our constitutional commitment to diversity compels us to *acknowledge* that there are other cultural contexts within broader South African society that will allow an anonymous-donor-conceived surrogacy child to flourish, and the nature and suitability of any given cultural context can be assessed on a case-by-case basis. Moreover, if we take the value of diversity seriously, we should *celebrate* the creation of novel types of family formation, like families formed by using a surrogate mother and where the child is conceived using male and female anonymous donor gametes.

In contrast, the tradition-based argument adopts a hegemonic approach that is anathema to the value of diversity. The tradition-based argument effectively seeks to enforce one set of cultural norms (in this case traditional black South African cultural norms) on all people of our country, hence abrogating the ‘right to be different’.

---


70 *Christian Education* (note 69 above) at para 24.


72 Ibid at para 19.

73 [2005] ZACC 19, 2006 (1) SA 524 (CC) (‘*Fourie*’).

74 Ibid at para 59.

75 Ibid at para 86.

76 The *AB* minority was alive to the value of diversity, and came to a conclusion about the family similar to the one I set out in this paragraph. See *AB CC* (note 4 above) at para 119 (It is constitutionally impermissible to say that families with children who are not genetically connected to their parents are significantly worse off. The Constitution instead celebrates this difference.)

77 Ibid at para 196 (Section 294 privileges one factor to the exclusion of all others.)
VIII  THE LAW SHOULD NOT GIVE EFFECT TO PREJUDICE

The final argument against the tradition-based argument that I propose is based on the idea that the law should not give effect to prejudice. This argument presents a paradigm shift away from the previous three counter-arguments. It is best introduced by recalling the case of *Van Rooyen*. It is safe to assume that homophobia existed in the white (likely Afrikaans) community in which the lesbian mother lived in the mid-1990s. As a result, her children could have been socially stigmatised and harassed because of her open homosexuality. Should this social reality have informed the court’s inquiry into the best interests of the child?

Although this question was not considered in *Van Rooyen*, it was answered in the affirmative in the American case of *Jacobson v Jacobson* – a case that predates *Van Rooyen* by about a decade. Similar to *Van Rooyen*, *Jacobson* was also a custody dispute between two divorced parents that centred on the mother’s new lesbian relationship. In *Jacobson*, the North Dakota Supreme Court awarded custody to the father, because it held that the children would ‘suffer from the slings and arrows of a disapproving society’ to a much greater extent than if they were to stay with the mother and her lesbian lover.79 The principle established by the *Jacobson* decision was that the court can consider prevailing societal prejudice when interpreting the content of the best interests of the child. If this principle were to be adopted in South African jurisprudence, how many children would be placed in the care of homosexual parents? If, in a specific case, it can be proven that the child is likely to be socially stigmatised and harassed if placed in the care of a homosexual parent, should this not require a balancing between the best interests of the child and the right to non-discrimination of the homosexual parent? Although South African case law has dealt with other aspects of homosexual parenthood, it has not yet dealt with this specific question.

In American case law, the *Jacobson* judgment was not the end of the road. Three years after the *Jacobson* judgment, the principle underlying this judgment, namely that it is acceptable to consider prevailing societal prejudice when interpreting the content of the best interests of the child, was rejected by the United States Supreme Court in *Palmore v Sidoti*.80 *Palmore* dealt with another type of societal prejudice different than that contemplated in *Jacobson*: racial prejudice. The facts were as follows: the mother and the father, both white, were divorced. Their child was placed in the custody of the mother. However, the following year the father filed suit to modify the custody order based on changed circumstances, being that the mother moved in with and subsequently married a black man. The Florida trial court ordered the child to be removed from the mother’s custody for the following reason:

This Court feels that despite the strides that have been made in bettering relations between the races in this country, it is inevitable that Melanie [the child] will, if allowed to remain in her present situation and attains school age, and thus more vulnerable to peer pressures, suffer from the social stigmatization that is sure to come.81

The Florida trial court’s judgment was affirmed by the Florida District Court of Appeal but was eventually reversed by the United States Supreme Court. The Supreme Court acknowledged the reality of racial prejudice in society, and that there was therefore a risk that the child may

---

78 314 NW 2d 78 (ND 1981).
79 Ibid at paras 81–82.
81 Ibid at 431.
A CRITIQUE OF AB V MINISTER OF SOCIAL DEVELOPMENT

suffer a variety of ‘pressures and stresses’ if she stayed with her mother and black stepfather. However, it held as follows in a unanimous ruling:

The question, however, is whether the reality of private biases and the possible injury they might inflict are permissible considerations for removal of an infant child from the custody of its natural mother. We have little difficulty in concluding that they are not. The Constitution cannot control such prejudices but neither can it tolerate them. *Private biases may be outside the reach of the law, but the law cannot, directly or indirectly, give them effect.*

The principle that the law should never give effect to prejudice has also crystallised in South African constitutional jurisprudence – although in another context, namely employment discrimination against persons with HIV. In *Hoffmann v South African Airways*, the Constitutional Court held that:

The constitutional right of the appellant not to be unfairly discriminated against cannot be determined by ill-informed public perception of persons with HIV. [...] Prejudice can never justify unfair discrimination. This country has recently emerged from institutionalised prejudice. Our law reports are replete with cases in which prejudice was taken into consideration in denying the rights that we now take for granted. Our constitutional democracy has ushered in a new era – it is an era characterised by respect for human dignity for all human beings. In this era, prejudice and stereotyping have no place. Indeed, if as a nation we are to achieve the goal of equality that we have fashioned in our Constitution we must never tolerate prejudice, either directly or indirectly.

This principle can find fruitful application to the tradition-based argument: first, the South African Constitution lists *social origin*, which includes clan or family membership, as a prohibited ground of discrimination. Second, the ‘slings and arrows of a disapproving society’ need not only be in the form of harassment or social stigmatising, but can also be through inculcating in the child’s psyche from a tender age the cultural precepts that are likely to cause the child to experience negative feelings about himself/herself based on his/her social origin. It may be subtler, but no less insidious. In this way, a child’s social environment uses the child’s own psyche as a weapon against the child’s self-identity and self-respect.

In this light, consider the first two premises of the tradition-based argument:

- In traditional black South African culture, (Value A), belonging to a clan is foundational for one’s identity, and (Value B) belonging to a clan depends on blood ties with the clan, or (in less rigid varieties of tradition) depends on knowing the child’s genetic origins. (Premise 1)
- Therefore (because of Premise 1), in traditional black South African culture, if a child was conceived using anonymous male and female donor gametes, (because of Value B) belonging to a clan would be impossible, and (because of Value A), the child will lack a foundational element of his/her identity. (Conclusion 1/Premise 2)

This constitutes prejudice against children based on their social origins – similar to prejudice based on inter alia race, sex, sexual orientation, and HIV status. Such prejudice may be beyond the reach of the law to control, but, as held in *Hoffmann*, in our constitutional dispensation...

---

82 Ibid at 433 (emphasis added).
83 *Hoffmann v South African Airways* [2000] ZACC 17, 2001 (1) SA 1 (CC) (‘*Hoffmann*’).
84 Ibid at paras 36–37, fn reference omitted.
86 Constitution s 9.
‘we must never tolerate prejudice’. By accepting societal prejudice as its very first premise, the tradition-based argument violates the principle that the law should never give effect to prejudice. Accordingly, the tradition-based argument is untenable in our constitutional dispensation. The tradition-based argument is in principle no different from arguing – based on the best interests and dignity of the child – that homosexual persons should not be allowed to have children based on persisting homophobia in certain parts of society, and that mixed-race couples should not be allowed to have children based on persisting racism in certain parts of society. Clearly, the law should not give effect to such prejudice.

The fact that social-origin prejudice is deeply rooted in cultural tradition offers no immunity for such prejudice. While s 31 of the Constitution protects the right of cultural communities to enjoy their culture, this right cannot be used to shield cultural practices that are inconsistent with the Bill of Rights; furthermore, as held in Fourie, ‘the antiquity of a prejudice is no reason for its survival’. It is worth remembering that the Constitution envisions South Africa to be an open society. Sachs J once remarked that ‘[t]he concept of an open society must indeed be regarded as one of the central features of the bill of rights’. The idea that tradition can never be sacrosanct is part of the very definition of an open society proposed by Popper, the philosopher who is most closely associated with the concept. He defined an open society as a society that ‘rejects the absolute authority of the merely established and the merely traditional while trying to preserve, to develop, and to establish traditions, old or new, that measure up to [the] standards of freedom, of humaneness, and of rational criticism’.

IX CONCLUSION

Earlier in this article, I posed the question: What deeply rooted beliefs could have caused the majority of the Constitutional Court in AB to negate the rule of law – by ignoring and even perverting the evidence – when it came to the issue of children’s blood ties with their parents? The likely answer, I suggest, is found in the sociological fact that traditional black South African cultural precepts about kinship formation, which place great emphasis on blood ties, continue to be influential in contemporary South African society. This fact can be used as the basis for a superficially convincing argument in support of the AB majority judgment – the tradition-based argument. Especially for people who themselves have been brought up and identify with traditional black South African cultural precepts about kinship formation, the tradition-based argument and its conclusion may seem self-evident. However, in this article I have endeavoured to show that when the tradition-based argument is subjected to careful analysis, it fails to convince. The most strident counter-argument is that the tradition-based argument relies on and fortifies cultural precepts that are discriminatory in nature, hence violating our Constitution’s commitment to equality. While the first three counter-arguments implicitly accept that it is constitutionally permissible to consider social prejudice when determining the best interests of the child, the fourth counter-argument strikes at the heart of the tradition-based argument by unmasking and confronting its basic assumption. In this light, the first three counter-arguments can be regarded as betrayals of the Constitution for implicitly accepting and

87 Hoffmann (note 83 above) at para 37.
89 Fourie (note 73 above) at para 74.
90 S v Lawrence; S v Negal; S v Solberg [1997] ZACC 11, 1997 (4) SA 1176 (Sachs J, concurring) at para 146.
hence legitimising the prejudiced basis of the tradition-based argument. The fourth counter-argument – the law should not give effect to prejudice – calls for a paradigm shift away from the way that both the *AB* majority and minority judgments approached the interpretation of the best interests of the child and dignity.

*AB* is a reminder that constitutional rights, in this case the best interests of the child and dignity, acquire their *meaning* through interpretation by the courts. Determining the meaning of a right inevitably requires value judgements that are informed by prevailing societal and cultural norms. This can be problematic, especially in contexts that are unconventional. Cultural norms may be deep-seated in society (or a section thereof), but not necessarily aligned with the values that the Constitution aspires to. The power of convention to subvert litigants and the courts from exercising proper (self-)scrutiny should not be underestimated. As a result of the blinding power of convention, rights that were intended to protect *against* prejudice, can paradoxically become *instruments* of prejudice.
The Best Interests of the Child
and the Constitutional Court

MEDA COUZENS

ABSTRACT: The Constitutional Court has developed a comprehensive child-friendly jurisprudence on the best interests of the child provision in s 28(2) of the Constitution. There are, however, concerns that the concept of the best interests of the child is being over-used by the Court to the detriment of other relevant children’s rights. The Court has not explicitly defined the content of s 28(2) in the name of preserving its flexibility. This article canvases the jurisprudence on the best interests of the child, and then it presents and analyses the use of s 28(2) by the Court in J v National Director of Public Prosecutions & Another (Childline South Africa & Others as amici curiae) and Raduwha v Minister of Safety and Security (Centre for Child Law as amici curiae). These cases show that despite the declared reluctance of the Court to give formal clarity to the content and the scope of s 28(2) of the Constitution, it may be starting to systematise its approach to the application of this section. In these cases, the Court spells out the legal content of the provision and uses it as a subsidiary tool in the absence of another legal provision relevant for the issue raised. These cases contribute to the clarification of the best interests jurisprudence. The further development of this good practice would be facilitated by the courts acknowledging the diversity of legal sources and functions of the best interests of the child concept. An awareness of the complex nature of the best interests concept enables its principled legal development, without endangering the flexibility of its application on which the success of the concept rests.

KEYWORDS: arrest, children’s rights, constitutional review, juvenile justice.

AUTHOR: Honorary Research Fellow and Member of the Navi Pillay Research Group, School of Law, University of KwaZulu-Natal, Durban. Email: couzensm@ukzn.ac.za

ACKNOWLEDGEMENTS: The author wishes to thank the anonymous reviewers and the editors of the journal for their constructive comments on various drafts of the article, and Associate Professor Ed Couzens for his assistance with editing this piece. Thanks are also due to the participants in the CCR IX workshop, Johannesburg, 3 August 2018, who provided feedback on an earlier version. The author takes responsibility for all remaining errors.
I INTRODUCTION

The Court has a rich children’s rights jurisprudence. A significant contribution is its comprehensive jurisprudence on the best interests of the child, for which the Court has been internationally praised.\(^1\) The constitutionalisation of the best interests of the child in s 28(2) of the Constitution of the Republic of South Africa, 1996 (‘Constitution’) has facilitated this development. However, the Court’s jurisprudence on the best interests provision is far from clear. It contains cases in which s 28(2) is applied either as a right, a principle, or a standard, without an explanation for the choice between the roles assigned to it. Arguably, this is not conducive to legal certainty as, presumably, different application techniques and outcomes accompany each of the mentioned roles. Further, the Court has been innovative in declaring that s 28(2) of the Constitution contains an independent right that applies discretely from the more specific rights of the child in s 28(1) of the Constitution. However, the Court has offered no explanation, and has explicitly refrained from assigning a fixed content to the right it identified in s 28(2). It reasoned that this would be contrary to the flexibility needed for this section to fulfil its purpose.\(^2\) This is problematic. A right has an ascertainable content that should be spelled out by courts which ought not shirk their duty to define that content. Failing to give content to a right runs the risk of negating the potential benefits which arise from its recognition as a right – its predictability, uniformity in application and certainty. In the case of s 28(2), this reluctance results in a failure to capitalise on the changed nature of the best interests of the child, and to put to rest criticisms in relation to its vagueness and indeterminacy.

This article argues that the Court needs to overcome its reluctance to define the content of s 28(2), viz, ‘A child’s best interests are of paramount importance in every matter concerning the child’. In spite of its reluctance, the Court has taken some unacknowledged steps in that direction, but its reasoning has not been sufficiently explicit and systematic. The article argues that the task of doing so will be easier if the Court acknowledges the complex functions that the best interests of the child requirement has come to fulfil and the changes in the nature of the concept that have arisen from its inclusion in human rights instruments such as the Constitution and the UN Convention on the Rights of the Child, 1989 (‘the CRC’). Two recent cases – \(J\) v National Director of Public Prosecutions & Another (Childline South Africa & Others as amici curiae)\(^3\) and \(Raduvha\) v Minister of Safety and Security (Centre for Child Law as amicus curiae)\(^4\) – develop the Court’s jurisprudence, and suggest directions in which the Court could take its case law so as to introduce some clarity in the judicial application of s 28(2) of the Constitution.

This contribution is structured as follows: part II contains a discussion about the changed nature of the concept of the best interests of the child in the era of human rights, including an


\(^3\) [2014] ZACC 13, 2014 (2) SACR 1 (CC) (‘\(J\) v NDPP’).

\(^4\) [2016] ZACC 24, 2016 (2) SACR 540 (CC) (‘\(Raduvha\)’).
introduction to the multifaceted nature of this concept. Part III provides a brief summary of the Court’s jurisprudence on the best interests, followed in part IV by a presentation of several arguments in favour of more judicial clarity in relation to the application of s 28(2). Part V, analyses *J v NDPP* and *Raduvha*, focussing on the contribution of these two cases in developing the Court’s jurisprudence on the independent application of s 28(2). Part VI contains this work’s conclusions.

II THE BEST INTERESTS OF THE CHILD IN THE ERA OF HUMAN RIGHTS

To understand the difficulties involved in defining the content of s 28(2) it is necessary briefly to present the changes which the concept has undergone as a result of being enshrined in a binding international instrument, the CRC. Article 3(1) of the CRC provides that –

> in all actions concerning children, whether undertaken by public or private social welfare institutions, courts of law, administrative authorities or legislative bodies, the best interests of the child shall be a primary consideration.⁵

A simple reading of the text suggests that its scope is wide, and includes ‘all actions concerning children’ taken by all branches of the state. This was an unprecedented enlargement of the scope of a concept most often encountered in private, family law litigation, or child protection proceedings concerning individual cases. The concept took off internationally, with the Committee on the Rights of the Child (‘the CRC Committee’) including art 3 amongst the four general principles of the CRC.⁶ Domestically, art 3 catalysed the extension of the standard of the best interests of the child in areas of law where this concept has not been applied before (such as juvenile justice and immigration) and to matters concerning children indirectly and not only directly.⁷

In 2013, the CRC Committee issued a general comment in which it interpreted art 3(1). It stated that this article contains a principle, a rule of procedure and an independent right.⁸ A striking (although not unusual for South African lawyers⁹) feature of the CRC Committee’s position is that art 3(1) contains an independent right despite the text not being formulated in rights language.¹⁰ It defined that substantive right as being –

> the right of the child to have his or her best interests assessed and taken as a primary consideration when different interests are being considered in order to reach a decision on the issue at stake, and the guarantee that this right will be implemented whenever a decision is to be made concerning a child, a group of identified or unidentified children or children in general.¹¹

---

⁵ African Charter on the Rights and Welfare of the Child, 1990/1999, art 4(1) reads: ‘In all actions concerning the child undertaken by any person or authority the best interests of the child shall be the primary consideration’.

⁶ CRC Committee *General Guidelines Regarding the Form and Content of Initial Reports to be Submitted by State Parties under Article 44, Paragraph 1(a), of the Convention* (1991) CRC/C/5 at para 13 (The other general principles are the rights to equality, survival and development, and participation).


⁸ CRC Committee *General Comment No. 14 (2013) on the Right of the Child to Have His or Her Best Interests Taken as a Primary Consideration art. 3, para. 1* CRC/C/GC/14 (‘General Comment 14’) at para 6.

⁹ Fitzpatrick (note 2 above).


¹¹ General Comment 14 (note 8 above) part I.A.
Article 3(1) has become one of the ‘stars’ of the CRC; and popular and successful as a litigation and advocacy tool. A few examples should suffice. In France, although courts are cautious in applying international treaties directly, art 3(1) is applied frequently, with significant positive consequences for children. It has been used in individual cases, but also as a supra-legislative provision capable of controlling the validity of legislation or secondary legislation. Article 3(1) has also been applied in other jurisdictions, both in ordinary and constitutional jurisprudence.

The European Court of Human Rights often engages with the concept of the best interests of the child, including by referring to art 3(1).

Despite its recognition in the CRC, the concept of the best interests of the child remains controversial. Concerns include its indeterminacy or vagueness; its potential to mask paternalistic decisions concerning children; and more recently, concerns about its over-use in argument and courts’ reasoning. A critical discourse on the use of the best interests of the child is therefore starting to emerge. Particularly relevant for this article is the view of Nigel Cantwell, who has criticised both the content of art 3(1) and its interpretation by the CRC Committee.

The thrust of Cantwell’s view is that ‘the prominent role now assigned to the “best interests of the child” is mistaken, even dangerous in a context where children have human rights’. The best interests was ‘a product of an era prior to children being explicitly granted human rights’, but the concept was nonetheless included in the CRC although children were to

---

13 M Couzens ‘France’ in Liefaard & Doek (note 10 above) at 123.
14 Ibid at 131.
15 Teoh (note 7 above); ZH (Tanzania) (FC) (Appellant) v Secretary of State for the Home Department (Respondent) [2011] UKSC 4; A, B, C and the Norwegian Association for Asylum Seekers (NOAS) v The State, Represented by the Immigration Appeals Board HR-2012-02399-P (case no. 2012/1042)(Norway Supreme Court).
20 Ibid at 62.
21 Ibid at 64.
have rights of their own. The extension of the scope of art 3(1) from specific matters to ‘all matters concerning children’ during the drafting of the CRC was not explained or justified and its implications were not assessed. Later, the CRC Committee unilaterally elevated art 3(1) to the rank of a general principle of the CRC, and embraced a ‘sacrosanct stance’ that the concept is of fundamental value for the CRC, without determining how and when its application could improve the protection of children’s rights. In its further developments of the concept in General Comment 14, the CRC Committee added to the confusion when it ‘invoked [it] pointlessly, that is, when reference to a right would or should suffice’. Cantwell argues that to recalibrate the approach to the best interests of the child, the concept should be used to fill gaps in the legal framework when ‘rights considerations alone do not provide sufficient guidance or grounds for decision-making’. Cantwell’s views are persuasive, but cogent views supportive of the position of the CRC Committee have also been expressed. It is beyond the scope of this article to delve into the debate, but two observations can be made.

First, the Cantwell criticism overlooks the possibility that the nature of what is monolithically called ‘the best interests of the child’ has changed under the influence of human rights standards. A new feature has been added to the concept: its traditional nature as a practical standard that enabled decision-makers to make decisions concerning individual children has been supplemented with human rights dimensions that strengthen it as a legal standard. Thus, what was essentially a checklist of factors (or a scale to weigh competing interests), has changed into a legal standard amenable to giving rise to entitlements and obligations. Early children’s rights writers, in similar vein to Cantwell, have argued that art 3(1) does not create specific obligations and entitlements. However, the law has developed in a different direction. More specifically, the transformation of the best interests of the child into a human rights standard requires that when determining what is in a child’s best interests, the rights of the child (i.e., not only the child’s welfare or what adults perceive to be the child’s welfare) be considered, and that a child capable of forming views has a right to be heard in all matters concerning the

---

22 Ibid at 65.
23 Ibid at 63.
24 CRC Committee General Guidelines (note 6 above).
25 Cantwell (note 19 above) at 64.
26 Ibid at 66.
27 Ibid at 69.
28 Sutherland & Barnes (note 16 above) (offers a more positive approach to the current development of the concept of the best interests of the child).
29 This transformation view can find support in treaty interpretation arguments such as a dynamic interpretation of the treaty (see Case Concerning the Dispute Regarding Navigational and Related Rights (Costa Rica v Nicaragua) 13 July 2009, ICJ para 64) or an interpretation in good faith of art 3(1) of the CRC (as per art 31(1) of the Vienna Convention on the Law of Treaties, 1969/1980), which does not render this article meaningless (Costa Rica v Nicaragua 2009 para 52) and which demands an integration of the best interests of the child concept into the rights-based reasoning endorsed by the CRC.
The weight to be given to the best interests of the child may also be different in that some human rights instruments contain a compromise between a lower weight to be given to them and expanding the ambit of the concept (to ‘all matters concerning a child’ – directly and indirectly, individually or collectively).

Second, putting aside the legal correctness of the current expansive approach to the best interests of the child, the reality remains that this perspective has been embraced in some jurisdictions, where it has had positive consequences for children individually and collectively. Thus, pragmatism surpassed the theoretical vulnerabilities of this concept. This raises questions for children’s rights researchers. For example, are concerns about the cogency of the current use of art 3(1) to be ignored because of the benefits that it can deliver? Or, can the best interests jurisprudence develop a sounder conventional legal foundation – meaning one that is unadventurous or one that conforms to mainstream legal discourse? Would the latter involve a decrease in the influence that the best interests of the child legal provisions have had as drivers of a special legal treatment for children? These questions cannot be fully addressed here, but the existing South African case-law suggests that the best interests of the child jurisprudence can develop more cogently. This requires an acknowledgement that the concept of the best interests of the child is complex and displays a multitude of dimensions. Distinguishing between these dimensions would enable some of the aspects of the concept to be defined, increasing legal certainty, while preserving the flexibility inherent in some other aspects of the concept.  

---

31 CRC Committee General Comment No 14 (note 8 above) at para 6; J v NDPP (note 3 above). Consideration of relevant rights and of the views expressed by a child result in the best interests of a child being determined based on some objective factors rather than on the subjective views and values of the decision-makers (T Boezaart ‘General Principles (ss 6–17)’ in C Davel & A Skelton (eds) Commentary on the Children’s Act (Revision Service 9, 2018; updated to August 2018) (Jutastat) at 8).

32 CRC, art 3(1) compare with African Charter, art 4(1).

33 CRC, art 3(1)(the best interests of the child are ‘a primary consideration’), and Constitution, s 28(2)(the best interests of the child are of ‘paramount importance’). The African Charter (art 4(2)) confers the highest weight, referring to the best interests of the child as ‘the primary consideration’.

34 This issue is developed in part VI below. Nonetheless, to understand better the arguments that follow, it is useful to mention that the legal dimensions of the best interests of the child (i.e. the in abstracto entitlements and obligations arising from s 28(2) of the Constitution) could be spelled out by a court, but what may be in the best interests of a child in a concrete dispute cannot be pre-determined in the abstract. As put by Boezaart, ‘[t]he question of precisely what a child’s best interests are is a factual question that has to be determined according to the circumstances of each individual case’ (Boezaart (note 31 above) at 6 (foonote omitted)).
III THE BEST INTERESTS OF THE CHILD IN SOUTH AFRICAN LAW:
A COMPLEX CONCEPT

The best interests of the child as a concept was used in South African law prior to the new constitutional dispensation, primarily to deal with child custody and the relationship between children and their parents, but also in relation to adoptions and child protection. With the advent of democracy, s 28 of the Constitution was drafted under the influence of the CRC. Section 28(2) reflects art 3(1) of the CRC to a significant extent, and it reads: ‘[a] child’s best interests are of paramount importance in every matter concerning the child’.

Like the CRC Committee, albeit independently, the Court employed a generous perspective in relation to s 28(2). In Fitzpatrick, the Court made the far-reaching pronouncement that the section contains a right independent of the more specific rights in s 28(1) of the Constitution, rejecting earlier views that it ‘is intended as a general guideline’. Thereafter, the position was reiterated in other cases, such as De Reuck v Director of Public Prosecutions (Witwatersrand Local Division) & Others, Sonderup v Tondelli & Another, M v S (Centre for Child Law as Amicus Curiae), and Director of Public Prosecutions, Transvaal v Minister of Justice and Constitutional Development & Others.

The Court has not explicitly articulated its approach to the best interests of the child despite the frequency with which it refers to s 28(2). Thus, in its case-law, s 28(2) is referred to

---


36 Mills (note 35 above) at 847–848.


38 Some differences exist in the textual formulation (note 33 above).

39 Fitzpatrick (note 2 above) at para 17.

40 Jooste v Botha 2000 (2) SA 199 (T) 210D.

41 [2003] ZACC 19, 2004 (1) SA 406 (CC) (‘De Reuck’).

42 [2000] ZACC 26, 2001 (1) SA 1171 (CC) (‘Sonderup’).

43 M v S (note 7 above).

44 DPP (note 2 above).

45 In this sense, its attitude is similar to that of the CRC Committee. Neither body has explained its far-reaching statements that the relevant best interests provisions contain an independent right. The reasons remain therefore a matter of speculation, creating concerns about the cogency of this approach.
alternatively as containing a ‘standard’, a ‘guiding principle’, or a ‘right’. The children’s rights literature has pointed out that in the jurisprudence of the Court the best interests of the child plays three functions: an interpretation tool for s 28(1) of the Constitution; a tool to establish the scope and potential limitations of other constitutional rights; and a right in itself.

It is apparent therefore that the best interests of the child plays a complex role. This complexity is compounded by s 28(2) being superimposed on the existing law, which already contained the principle of the best interests of the child, and by the utilisation of the best interests concept in child law legislation: Children’s Act 38 of 2005 (‘the Children’s Act’) and the Child Justice Act 75 of 2008 (‘the Child Justice Act’). Although the Constitution, the common law and statutes may all operate with nominally the same concept – ‘the best interests of the child’ – its purpose and meaning may differ. In terms of the common law, the best interests of the child has a limited ambit, and concerns the relationship between parents and children. It focuses on the individual child, is outcome-oriented (it seeks to deliver the best outcome for the child by giving priority to his/her interests), and when applied it determines the outcome.

The constitutional best interest differs. It applies not only to decisions made by high courts as upper guardians of all children but to ‘every matter concerning the child’, and to children individually and collectively. Statutory provisions held to be inconsistent with s 28(2) of the Constitution have been invalidated, arguably the most far-reaching constitutional remedy. Importantly, s 28(2) of the Constitution was declared by the Court on several occasions to be subject to limitations under s 36. This interpretation is clearly different from the common law best interests concept. Under the common law, the best interests analysis is not subject to such structured limitations.

Statutes add to the complexity of the concept. Section 6(2)(a) of the Children’s Act requires that in all proceedings concerning a child, his/her best interests must be respected, protected, promoted and fulfilled. Section 7 provides that when the Act requires the application of the best interests of the child, the factors in s 7(1) are to be considered. The factors referred to in this section are to be applied when the Act itself requires the application of the best interests of the child, and are not mandatory in cases beyond the ambit of the Act. In any case, most of the factors mentioned in s 7(1) would be difficult to apply beyond matters governed by the

---

46 Fitzpatrick (note 2 above) at para 18 (best interests of the child is referred to as a ‘standard’); Fitzpatrick (note 2 above) at para 17 (best interest of the child described as an independent ‘right’).

47 M v S (note 7 above) at para 22.

48 Fitzpatrick (note 2 above) at para 17; M v S (note 7 above) at para 22. Teddy Bear Clinic for Abused Children & Another v Minister of Justice and Constitutional Development & Another [2013] ZACC 35, 2014 (2) SA 168 (CC) (‘Teddy Bear’) used all the above terms; see, for example, para 79 where the Court finds the impugned legislation contrary, inter alia, to the ‘best-interests principle’, only to say a few lines later that ‘the right in terms of section 28(2)’ can be limited under s 36 (my emphasis).


50 Ibid at 40–47.

51 Fitzpatrick (note 2 above); J v NDPP (note 3 above); and C & Others v Department of Health and Social Development, Gauteng & Others [2012] ZACC 1, 2012 (2) SA 208 (CC) (‘C v Department of Health, Gauteng’).

52 Fitzpatrick (note 2 above) at para 20; Sonderup (note 42 above) at para 29; De Reuck (note 41 above) at para 55; and J v NDPP (note 3 above).

53 The standard applied in limited circumstances which involved people with special responsibilities in relation to the child (i.e. the parents).
Act. Further, s 9 states that ‘[i]n all matters concerning the care, protection and well-being of a child the standard that the child’s best interests is of paramount importance, must be applied’. The later provision is similar to s 28(2) of the Constitution by giving ‘paramount importance’ to the best interests of the child, but has a more limited scope in that it concerns only matters concerning the care, protection and well-being of children. There are, however, provisions which make the best interests of the child the determining factor for the courts, such as in the context of adoption, HIV testing, etc. The Child Justice Act also refers to the best interests of the child in several sections, and implies that in proceedings involving children ‘the best interests of the child are at all times of paramount importance’. The ‘paramount importance’ of the best interests of the child is mentioned only once (above), and in other instances, the best interests of the child is approached as one of the factors potentially relevant.

This brief presentation shows the variety of meanings and roles compressed into a single expression: ‘the best interests of the child’. Common law, statutory and constitutional approaches to the best interests coexist, but there is no perfect overlap between their scope, application and effects. Different legal sources deal with different aspects of the best interests of the child in different ways, but none defines the best interests or exhausts its meaning and its legal consequences. While the aim of securing the best possible outcome for the child is central to all approaches, the routes taken are different. Considering this complexity, it is difficult for the Court to invoke the need to preserve the flexibility of the best interests of the child in order to refrain from giving more structure to its approach to this concept. Its position is particularly problematic when it concerns the far-reaching and novel statement that s 28(2) contains an independent right.

It is not certain why the Court declared that s 28(2) contains an independent right. The is not formulated in rights language, and declining to characterise it as a right would not have denied it legal effect. Legal norms are not deprived of normative force just because they do not contain substantive rights or because they have a more general formulation than other norms. The Court’s own jurisprudence shows that constitutional norms and principles – including provisions in the Bill of Rights, the rule of law or separation of powers – are

54 A similar approach is to be found in Children’s Act, s 64(1)(a)(‘[T]aking into account the best interests’ of the child).
55 Children’s Act, s 230(1)(a).
56 Children’s Act, s 130(1).
57 Children’s Act, ss 29(1)(3), 55(1), 61(1)(c), 116(2), 151(8).
58 Child Justice Act, ss 9(1)(b), 24(3)(a), 30(3)(a).
59 Child Justice Act, s 80(1)(d).
60 DPP (note 2 above) para 190 (Court mentions its upper guardian functions alongside s 28(2)). Also Van der Burg & Another v National Director of Public Prosecutions & Another [2012] ZACC 12, 2012 (2) SACR 331 (CC)(‘Van der Burg’) para 68 (‘The High Court is not only the upper guardian of children, but is also obliged to uphold the rights and values of the Constitution’).
61 Bonthuys (note 18 above) at 27 (remarking that the Court has seldom treated s 28(2) as an independent right in that it has not defined its content and it has seldom utilised s 36 to justify limitations to the best interests). However, the Court has considered possible limitations to s 28(2) in terms of s 36 of the Constitution. C v Department of Health, Gauteng (note 51 above) at paras 80–83; Teddy Bear (note 48 above) at para 79; J v NDPP at para 46.
63 Glenister v President of the Republic of South Africa & Others [2011] ZACC 6, 2011 (3) SA 347 (CC) (Impugned legislation was found contrary, amongst others to s 7(2), which is not a substantive provision in the Bill of Rights).
judicially enforceable despite not being declared rights and having wide normative spans.\textsuperscript{64} International writers\textsuperscript{65} and foreign case-law\textsuperscript{66} also support the view that abstract norms are judicially enforceable.

To the knowledge of the current writer, no author, court\textsuperscript{67} or international body has made similar statements prior to that of the Court in \textit{Fitzpatrick}. The Court itself has only explicitly identified the right in \textit{DPP}, where it referred to it as ‘the right to have the child’s best interests given paramount importance in matters concerning the child’.\textsuperscript{68} Until the CRC Committee embraced the same approach in 2013, there had been no source other than the Court able to clarify this new take on the best interests of the child. Regrettably, the Court failed to do so, although it has continued to apply s 28(2) in ways which have clearly advanced children’s interests.

While the Court has been reluctant to assign content to s 28(2), including the right component of the provision, a close analysis of its case-law gives an indication of the obligations and entitlements which the Court associates with this section. This contribution does not inquire into whether this content is commensurate to that of a right or whether it can be equally assigned to a constitutional norm of a different nature. Addressing this question requires a different approach, which is beyond the scope of this article.\textsuperscript{69} Nonetheless, several normative themes can be identified in the Court’s application of s 28(2),\textsuperscript{70} albeit that it is not certain whether they are a part of the content of the independent right identified by the Court, or of the wider normative content of this provision, which includes its functioning as a guiding principle or constitutional standard.

A useful departure point is the unpacking of Sachs J’s view on s 28(2) in \textit{M v S} by Gallinetti, who distils the following requirements from the decision:

\textsuperscript{64} Examples include the rule of law (\textit{Fedsure Life Assurance Ltd v Greater Johannesburg Transitional Metropolitan Council} 1998(2) SA 374 (CC)), the separation of powers (\textit{South African Association of Personal Injury Lawyers v Heath \\& Others} [2000] ZACC 22, 2001 (1) SA 883 (CC)). In environmental law, the National Environment Management Act 107 of 1998, s 32(1) makes the principles of the Act independently justiciable, without them containing rights.

\textsuperscript{65} Conforti argued that norms with a general or indeterminate formulation cannot be denied judicial application ‘especially when they contain declarations of principles rather than specific rules’, because legal principles are capable of judicial application. B Conforti \textit{International Law and the Role of Domestic Legal Systems} (1993) 28–29.

\textsuperscript{66} \textit{A, B, C and the Norwegian Association for Asylum Seekers} per Justice Bårdsen (joined in dissent by four other judges), who argues that the normative force and the justiciability of a legal norm (CRC, art 3(1) in that case) are not erased by its generality (paras 116–120). It is interesting to note that, like the South African Constitutional Court, this judge would have issued a judicial declaration of unlawfulness in relation to CRC, art 3(1) alone (incorporated verbatim in the Norwegian law) and would have provided a judicial remedy.

\textsuperscript{67} Gaudron J in \textit{Teoh} (note 7 above) said that ‘it is arguable that citizenship carries with it a common law right on the part of children and their parents to have a child’s best interests taken into account …’ (para 4). Unlike in South Africa, this thesis that the consideration of the best interests of the child constitutes a right of sorts has not taken off the ground.

\textsuperscript{68} \textit{DPP} (note 2 above) at para 73.

\textsuperscript{69} Arguably, the clarification of this issue requires a theoretical investigation into the meaning of the term ‘right’, followed by an assessment of how its judicial enforceability differs from that of legal norms not containing a right.

First, consideration of the interests of children; second, the retention in the inquiry of any competing interests because the best interests principle does not trump all other rights; finally, the apportionment of appropriate weight to the interests of the child.\(^{71}\)

Two obligations seem to arise from s 28(2): to consider (i.e. take into account) the interests of children, and to give ‘appropriate weight … in each case to a consideration to which the law attaches the highest value, namely, the interests of children who may be concerned’.\(^{72}\) This approach is supported by the Court’s formulation in *DPP*: ‘the right to have the child’s best interests given paramount importance in matters concerning the child’.\(^{73}\)

These obligations can be unpacked further. The obligation to consider the interests of the child requires a court to be informed about the impact of its decision on children. This may be achieved through the appointment of a curator *ad litem*,\(^{74}\) or information being provided by relevant court officers,\(^{75}\) or by listening to children and their parents.\(^{76}\) In cases where the parents are subject to law enforcement actions by the state, the interests of the children are to be assessed independently of those of their parents\(^{77}\) and the state has to minimize the harmful consequences of its interference with the family environment.\(^{78}\) Section 28(2) requires consideration of children’s interests in matters involving children or only affecting them. At times, a court may need to consider the interests of children *ex officio* if the parents or those supposed to safeguard the interests of the children fail to do so.\(^{79}\)

Giving ‘appropriate weight’ to the best interests of the child does not mean that such interests trump all other legitimate interests. The Court has been clear that s 28(2) can be limited according to s 36 of the Constitution.\(^{80}\) Section 28(2) requires that individual best interests be safeguarded, including by enabling the delivery of child-centred remedies.\(^{81}\) The law must therefore create conditions for decision-makers to respond to individual situations.\(^{82}\)

The above entitlements and obligations based on 28(2) arise from an academic analysis of judgments, which ultimately cannot be a substitute for binding judicial pronouncement, which in my view is currently needed.

---


\(^{72}\) *M v S* (note 7 above) at para 42.

\(^{73}\) *DPP* (note 2 above) at para 73 (per Ngcobo J).

\(^{74}\) In *Van der Burg* (note 60 above), the *amicus* relied on s 28(2) to request the appointment of a curator *ad litem*, while in *Du Toit & Another v Minister of Welfare and Population Development & Others (Lesbian and Gay Equality Project as Amicus Curiae)* 2003 (2) SA 198 (CC) at para 3, the appointment of the curator *ad litem* was linked to s 28(1)(h).

\(^{75}\) *Van der Burg* (note 60 above) at para 68; *M v S* (note 7 above) at para 36.

\(^{76}\) *C v Department of Health, Gauteng* (note 51 above) at para 77.

\(^{77}\) *M v S* (note 7 above) at para 18; *Van der Burg* (note 60 above) at para 68, 70.

\(^{78}\) *M v S* (note 7 above) at para 42.

\(^{79}\) *Van der Burg* (note 60 above) at para 68.

\(^{80}\) See note 51 above.

\(^{81}\) In *Sonderup*, the Court ordered the immediate return of a child abducted by her mother from the country of habitual residence, but held that ‘[p]ursuant to s 38, read with s 28(2), this Court is entitled to impose conditions [to her immediate return] in the best interests of Sofia.’ *Sonderup* (note 42 above) at para 51. See also *Head of Department, Department of Education, Free State Province and Welkom High School & Others* [2013] ZACC 25, 2014 (2) SA 228 (CC) at para 119 (Although a violation of children’s rights was not established, the Court granted a pre-emptive remedy because, otherwise, children would have been exposed to the risk of having their rights violated by school policies that they are unlikely to challenge).

\(^{82}\) *Fitzpatrick* (note 2 above) at para 20; *M v S* (note 7 above) at para 24.
IV  THE NEED FOR MORE JUDICIAL CLARITY IN RELATION TO THE APPLICATION OF S 28(2)

There are several pressing reasons why more clarity is needed on the application of s 28(2), especially as an independent right. These include preserving legal certainty; increasing the legitimacy of the powerful position currently occupied by s 28(2) by justifying the identification therein of a right; preserving the integrity of s 28(2) by ensuring a transparent application of s 36; and the divergent views taken by judges in relation to the application of s 28(2). In this part, I elaborate on these reasons.

First, clarity is needed in order to preserve legal certainty. When a concept plays as many roles and can have as many distinct consequences as the best interests of the child, there should be clarity on how to use the concept (i.e. as an independent right, a guiding principle, a standard, an interpretation tool, or a procedural rule).

Second, declaring that s 28(2) contains an independent right goes beyond the express wording of the section. Apart from raising questions about the legitimacy of this approach, ‘reading into’ the Constitution a right which has not been spelled out by the drafters extends considerably the sphere of rights for children. While this is no doubt positive for them, it may have an impact on the rights of others and the obligations of the state. In Fitzpatrick, the Court stated:

Section 28(1) is not exhaustive of children’s rights. Section 28(2) requires that a child’s best interests have paramount importance in every matter concerning the child. The plain meaning of the words clearly indicates that the reach of section 28(2) cannot be limited to the rights enumerated in section 28(1) and section 28(2) must be interpreted to extend beyond those provisions. It creates a right that is independent of those specified in section 28(1).

Thus, in the view of the Court, the interests of children, to which paramount consideration needs to be given under s 28(2), extend beyond those that are given explicit constitutional protection in s 28(1). This is a far-reaching statement, which results in a constitutionalisation of interests that have not been explicitly protected by the drafters of the Constitution. While it provides a positive counterweight to possible legislative oversight in relation to children’s interests, it also raises questions as to whether courts may not be allocating themselves Constitution-making powers.

An illustration is provided by the Teddy Bear case, where the Court found statutory provisions criminalising certain consensual sexual acts between teenagers contrary to, among others, s 28(2). The reasoning of the Court was that legislation exposed children to harm consisting of the ‘negative impact’ of the contact with the criminal law system. A general protection of children from harm is not explicitly provided for in the Constitution. Children have the right to be protected against serious forms of harm – ‘maltreatment, neglect, abuse or

83 Fitzpatrick (note 2 above) at para 17.
84 Criminal Law (Sexual Offences and Related Matters) Amendment Act 32 of 2007, ss 15 and 16 were declared constitutionally invalid to the extent that they imposed criminal liability on children below the age of 16.
85 Teddy Bear (note 48 above) at para 71. In the same paragraph, the Court noted: ‘The best-interests principle also applies in circumstances where a statutory provision is shown to be against the best interests of children in general, for whatever reason. As a matter of logic what is bad for all children will be bad for one child in a particular case’. Essentially, this seems to say that what is bad for children is contrary to s 28(2). This is a very wide interpretation of the scope of this section that, left unqualified, risks opening the door to paternalistic arguments which define the best interests of the child from an unfettered adult perspective. It might be that the Court deliberately gave s 28(2) a broad construction given that potential limitations may be justified under s 36.
degradation\textsuperscript{86} – but contact with the criminal justice system, without more, does not amount to such prohibited treatment despite being harmful. In the absence of further elaboration, the position of the Court in \emph{Teddy Bear} suggests that any harm done to a child/children is a breach of s 28(2) of the Constitution, regardless of whether the harm if grave or not. The consequence of such an approach can be significant, considering the obligation of the courts to invalidate legislation or official conduct inconsistent with the Constitution.

A third reason why it is important to know what the content of the right in s 28(2) is, and when such a right applies, is that should a limitation of that right be pleaded,\textsuperscript{87} it has to meet the stringent requirements in s 36 in order to be defensible. This is a positive feature of the jurisprudence. The concerns in relation to the best interests of the child include the alleged arbitrariness in decision-making and the manipulation of the standard to suit the interests of adults. Reliance on s 36 to justify the limitation of s 28(2) mitigates such risks, introduces structure and reduces the potential arbitrariness in the process of limiting the best interests of the child. Nonetheless, the application of s 36 is difficult if it is not clear what the right allegedly limited requires. Further, given the procedural nature of the right in s 28(2), as formulated by the Court – imposing an obligation to consider and to give appropriate weight to the best interests of the child – one may question whether a limitation to that right can ever meet the criteria in s 36. What can justify a limitation of a right to give primary consideration to the best interests of the child, which implies no automatic commitment to an outcome which favours the child? What does a justifiable limitation to this right mean: giving no consideration at all to the best interests of the child; or not prioritising the interests of the child? If the former, can there be any reasonable and justifiable reason in a society based on dignity, equality and freedom which justifies the interests of children not being seriously considered and weighed against other legitimate interests? If the latter, is it necessary to engage with s 36 considering that s 28(2) does not require that the interests of children be always prioritised? When the Court used s 36 to justify limitations to s 28(2) it arguably envisaged the common law best interests standard and its prioritisation of children’s interests, rather than the constitutional standard, which does not require that such priority be automatically given.\textsuperscript{88}

Fourth, certain cracks seem to be appearing in the judicial application of s 28(2), which show that a concept whose application was once uncontroversial may have lost its unifying power,\textsuperscript{89} with some judges seemingly developing a more circumspect attitude in relation to its application. Arguably, the uncertainty about the scope, role and consequences of s 28(2) has contributed to this. The Court has systematically avoided defining the legal content of s 28(2)

\begin{itemize}
  \item \textsuperscript{86} Constitution, s 28(1)(d).
  \item \textsuperscript{87} It is only rights whose limitations must comply with s 36 requirements, and not the limitations in other legal standards (such as guidelines or principles).
  \item \textsuperscript{88}\textsuperscript{88} Sonderup (note 42 above)(The Court relied on s 36 to justify why it did not prioritise the short term interests of the child). In \emph{Teddy Bear} (note 48 above) or \emph{J v NDPP} (note 3 above), it relied on s 36 to establish if there were any good reasons for the law not to prioritise a best outcome for children. Although these judgments are child rights-friendly, the problem remains that of the chameleonic use of s 28(2) by the Court, which starts by engaging with the constitutional dimension of the best interests of the child but disposes of the legal matter using the common law, outcome-oriented dimension of the concept.
  \item \textsuperscript{89} In certain cases in which s 28(2) has played a major role, there was no dissent on the application of this section and its effects (\emph{Fitzpatrick} (note 2 above); Sonderup (note 42 above)‘ \emph{De Reuck} (note 41 above); \emph{DPP} (note 2 above)).
\end{itemize}
(see, for example, Fitzpatrick\textsuperscript{90} and DPP\textsuperscript{91}) in the name of preserving its flexibility (on which the strength of the concept was said to rest).\textsuperscript{92} This creates a double danger: a temptation to over-use the best interest of the child given its normative elasticity, which could result in its normative bloating to the detriment of applying and developing the more specific rights of the child; and the risk of being overlooked by the courts given its uncertain scope and legal consequences.

Some clarification is needed on these points. Academics have expressed concern that judges may rely too easily on s 28(2), without making an effort to construct their reasoning on the more precise requirements of s 28(1) provisions.\textsuperscript{93} Some cases have exposed the vulnerability of this approach, possibly to the detriment of advancing the rights of children. In C v Department of Health, Gauteng, the impugned provisions of the Children’s Act were declared constitutionally invalid because they were found to breach s 28(2) (and s 34). Writing for the majority, Yacoob J did not engage with ss 28(1)(b) and (d); which were nonetheless considered by Skweyiya J (with Froneman J; concurring) and Jafta J (with Mogoeng CJ; dissenting). In Le Roux & Others v Dey; Freedom of Expression Institute & Another as Amici Curiae\textsuperscript{94} (‘Le Roux’) Yacoob J (dissenting), invoked s 28(2) to justify his child-focused approach to the law of defamation. Only Skweyiya J rallied to his reasoning. The disconcerting aspect in Le Roux is that, after years of best interests case-law, the majority did not even mention s 28(2) in its reasoning. It is possible therefore that when a legally-diffuse standard such as the best interests of the child comes face-to-face with well-established, hard law (law of delict in this case) some judges may be reluctant to consider it.

Recent Supreme Court of Appeal (SCA) cases add to this concern. In Centre for Child Law & Others v Media 24 Limited & Others\textsuperscript{95} (‘Centre for Child Law v Media 24’) the majority acknowledged the reliance on s 28(2) by the appellants but then ignored this section in its reasoning. It rejected the appellants’ claim to extend the anonymity of the victims after they turned 18, but accepted the claim for victim extension under s 9 of the Constitution,\textsuperscript{96} and not s 28(2). No explanation is given for snubbing the latter. Willis JA for the minority (joined by Mocumie JA) relied heavily on s 28(2), which he referred to as ‘the animating principle’\textsuperscript{97}

\textsuperscript{90} Fitzpatrick (note 2 above) at para 18, footnotes omitted (The Court said that ‘the ‘best interests’ standard appropriately has never been given exhaustive content in either South African law or in comparative international or foreign law. It is necessary that the standard should be flexible as individual circumstances will determine which factors secure the best interests of a particular child).

\textsuperscript{91} DPP (note 2 above) at para 73 (‘It is neither necessary nor desirable to define with any precision the content of the right to have the child’s best interests given paramount importance in matters concerning the child’).

\textsuperscript{92} AB v Pridwin Preparatory School [2018] ZASCA 150, 2019 (1) SA 327 (SCA) at para 30 (‘Pridwin Preparatory School’)(‘It is unnecessary to determine the content of this right because it provides an adequate benchmark for the treatment and protection of children in its present form.’). If the highest court in the land is of the view that it is not necessary to determine the content of the right in s 28(2), then lower courts are entitled to follow its lead.

\textsuperscript{93} Bonthuys (note 18 above); Couzens (note 18 above); Skelton (note 18 above); M Couzens ‘The Constitutional Court Consolidates its Child-focused Jurisprudence: The case of C v Department of Health and Social Development, Gauteng’ (2013) 130 South African Law Journal 672.

\textsuperscript{94} [2011] ZACC 4, 2011 (3) SA 274 (CC).

\textsuperscript{95} [2018] ZASCA 140, [2018] 4 All SA 615 (SCA). The case concerned the constitutional validity of s 154(3) of the Criminal Procedure Act 51 of 1977, which failed to protect the anonymity of children as victims of crimes in criminal proceedings. In addition to extending the protection of anonymity to child victims (the victim extensions), the applicants sought the preservation of anonymity of victims beyond the age of 18, except when decided otherwise by a court (the adult extension).

\textsuperscript{96} Ibid at para 29.

\textsuperscript{97} Ibid at para 64.
of the case. While he relied on s 28(2) to justify the victim extension, the adult extension was justified under the rights to dignity and privacy (rather than s 28(2)).

The explanation given by Willis JA as to the distinction between the majority and minority judgments, which he considered to be separated by a ‘philosophical ocean’, is interesting. This comment alerts us to the fact that the consideration of s 28(2) may be viewed by some judges as a matter of philosophical choice rather than as a matter of applying a relevant and binding legal standard.

Pridwin Preparatory School concerned a challenge by the parents of two children enrolled in a private school to the school’s decision to terminate the contract between it and the parents because of the parents’ misconduct. The parents argued that by terminating the contract, the school breached ss 28(2) and 29(1)(a) of the Constitution. Cachalia JA for the majority considered s 28(2), but decided that it did not create a right for the parents (or the children) to be heard by the school before the contract was terminated, as argued by the parents. Mocumie JA dissented, and found that the contractual clause which gave the school the right to terminate the contract without hearing the children was unconstitutional and that the termination of the contract was unfair because the views of the two children were not considered and given appropriate weight. The judgment by Mocumie JA pays close attention to s 28(2) and its implications. Without being error-free, her judgment questioned the method followed by the school to assess the best interests of the two expelled children and expressed the view that contracts between private schools and parents are not typical commercial contracts, and they may require that the courts develop the law of contract.

The cases show that best interests arguments do not always enjoy full support from judges or lead to a uniform application of s 28(2). Arguably, absence of clarity in relation to how and when s 28(2) applies may lead to some judges avoiding this standard in favour of more certain, less amorphous norms. It comes as no surprise therefore that when legal reasoning based on s 28(2) is advanced in areas of law not accustomed to child-related concerns, judges are criticised for being ‘misdirected by … unwarranted bias towards the children’s rights’. An approach to the best interests of the child presented in a conventional legal algorithm may be more appealing for courts and may improve the quality and the consistency of the application of s 28(2). This section has become so important for the children’s rights jurisprudence that continuing uncertainty may undermine the benefits it is able to deliver.

The jurisprudence is nonetheless evolving, and the two cases discussed below show the benefits which arise from the Court’s position that s 28(2) contains a legal norm which can

---

98 Centre for Child Law v Media 24 (note 95 above) at paras 74 and 83 (per Willis JA).
99 Ibid at para 97 (‘In my opinion, when it comes to the disclosure of the identity of childhood victims of crime, logic, common sense and ordinary, everyday morality generate a constitutional imperative.’).
100 Pridwin Preparatory School (note 92 above).
101 Ibid at para 39 (Cachalia JA found that s 29(1)(a) did not apply, because the school was not performing a constitutional function (i.e. provision of basic education)).
102 The Court held that the cases relied on by the parents do not create a general right to be heard and were not relevant for the termination of a private contract. Ibid at paras 33, 37. The parents had relied upon C v Department of Health, Gauteng (note 51 above) and J v NDPP (note 3 above).
103 C v Department of Health, Gauteng (note 51 above) is not a ‘medical treatment’ case (at para 96), and s 31(2) of the Constitution (relied on in para 124) is not relevant to the case.
104 Pridwin Preparatory School (note 92 above) at para 124. An appeal by the parents was heard by the Court on 16 May 2019, and judgment was reserved.
be applied independently. This is not to say that in these cases the Court has necessarily applied s 28(2) as a right. What the Court has nonetheless done, in unanimous judgments, has been to cement the recognition of the independent normative force of s 28(2) further and, in the process, offer some guidance on when this section can be applied independently in a predictable and principled way.

V  J v NDPP AND RADUVHA: RECENT POSITIVE EXAMPLES OF THE APPLICATION OF S 28(2)

In this part, I provide a brief account of J v NDPP and Raduvha, followed by a discussion of their significance for developing the jurisprudence of the Court on the independent application of s 28(2).

A  J v NDPP: Overcoming concerns about identifying the content of s 28(2)?

This case concerned the constitutional validity of s 50(2) of the Criminal Law (Sexual Offences and Related Matters) Amendment Act 32 of 2007, which provided that upon sentencing a person for a sexual offence against a child or a disabled person, a court must order the entering of the particulars of that offender in a National Register for Sex Offenders. Several adverse consequences arose from such registration, and in certain cases, including that of the applicant, the particulars could never be removed. The applicant, a 14-year-old child at the time of the offences, was sentenced on several counts of rape committed against younger children. The sentencing magistrate made an order for the child's details to be entered into the Register. When the matter came before the High Court, the Court decided, inter alia, that the rights of the child offender had been violated by the above section, which was declared unconstitutional.

In confirmation proceedings, the conflict between the impugned text and the Constitution was argued on different bases by the parties and the amici. Thus, the written arguments of the applicant relied on the rights to dignity, privacy, fair labour practices, and freedom of trade, occupation and profession; in oral argument, the applicant relied on s 34 of the Constitution; while the amici relied on s 28(2) of the Constitution. The respondents conceded conflict with s 35 of the Constitution. Skweyiya ADCJ (as he then was) considered as ‘correct’ the position of the amici that ‘the starting point for matters concerning the child is s 28(2)’. He found that the challenged statutory provision was contrary to s 28(2) of the Constitution, which made it unnecessary to consider the other grounds invoked by the parties.

To decide that the mandatory entry into the registry was contrary to the best interests of the child, the Court established that s 28(2) required that the law should generally provide for a differentiation between adult and child offenders; that the law should make allowance for individualised treatment of the child and a consideration of the representations made by

106 J v NDPP (note 3 above) at paras 21–25.
107 Ibid at para 25.
109 Ibid at para 33.
110 Ibid at para 34.
111 Ibid at para 35.
112 Ibid at para 45.
the child throughout the criminal justice process.\textsuperscript{113} The mandatory registration of the child offender was inconsistent with these requirements, which required that the Court assess their justification under s 36.\textsuperscript{114} The Court held that the limitation could not be justified.

\textit{J v NDPP} is perhaps the most streamlined judgment in which the Court engaged with s 28(2) as the sole reason for a judgment and as an independent right. From the multitude of legal grounds invoked by the parties, the Court decided to base its reasoning on s 28(2), which had been raised by the \textit{amici}. Beyond the general statement that this is a ‘starting point’ in matters concerning children, no other explanation for this choice is provided by the Court. Nonetheless, departing from its reluctance to define the content of s 28(2) as expressed in \textit{Fitzpatrick} and \textit{DPP}, the Court clearly identified three legal requirements arising from the provision. The Court also moved beyond the primarily procedural approach to s 28(2) espoused in \textit{M v S}, and derived a new layer of specific entitlements in the context of juvenile justice (special and individualised treatment for children, and listening to child representations).

\textbf{B Raduvha: The emergence of a subsidiarity approach to the independent application of s 28(2)?}

Ms Raduvha, the applicant, instituted a claim against the Minister of Safety and Security for damages arising from an alleged wrongful arrest and detention when she was 15 years old. The case raised two overall issues for the Court: firstly, the meaning of the best interests of the child and these interests being accorded paramount importance; and secondly, the impact of these best interests on the duty of police officers to arrest under s 40 of the Criminal Procedure Act 51 of 1977 (‘CPA’).\textsuperscript{115}

The incident which led to Ms Raduvha’s arrest occurred on 6 April 2008 when two police officers were sent to her home to investigate a complaint regarding the applicant’s mother. When the police officers attempted to arrest her mother, Ms Raduvha interposed herself between them and her mother. This was regarded by the officers as being an unlawful obstruction of their lawful duties, for which they arrested the applicant based on s 40(1)(j) of the CPA. The applicant’s mother was also arrested. They were detained at a police station for 19 hours and thereafter released on warning. The prosecutor refused to prosecute them. Ms Raduvha was unsuccessful in her claim for damages in the High Court; and the case reached the Court on appeal.

Ms Raduvha argued that the officers acted unlawfully and arbitrarily in arresting her, in that although s 40(1) of the CPA authorised them to arrest her, it provided the police with a discretion on whether or not to do so, which discretion the officers failed to exercise.\textsuperscript{116} The applicant also argued that the officers failed to consider and accord paramount importance to her best interests, and thus did not give effect to s 28(2) of the Constitution.\textsuperscript{117} The detention of the applicant was challenged under s 28(1)(g) of the Constitution, with the argument that it was not a measure of last resort since she could have been left in the care of her father.

\begin{thebibliography}{9}
\bibitem{113} Ibid at para 42.
\bibitem{114} Ibid at para 46.
\bibitem{115} \textit{Raduvha} (note 4 above) at para 5.
\bibitem{116} Ibid at para 16.
\bibitem{117} Ibid at para 17.
\end{thebibliography}
Amongst others, the case raised the lawfulness of the applicant’s arrest and detention.\footnote{\textit{Ibid} at para 28. The other two issues concerned whether s 28(2) of the Constitution created an additional jurisdictional requirement for a lawful arrest under s 40(1) of the CPA, and whether the Court should establish the damages.} Important for the purposes of this article is the identification by the Court of the constitutional standard against which the lawfulness of the arrest was to be decided. The issue arose because s 28(1)(g) of the Constitution deals explicitly only with the detention of children in conflict with the law and not with their arrest. Relying on the wording of s 35 of the Constitution, the Court decided that arrest and detention are two different processes,\footnote{\textit{Ibid} at para 36.} and thus the ‘last resort’ requirement in s 28(1)(g) did not apply to the arrest of the child.

Alternative grounds had to be used to assess the lawfulness of the arrest. Two such grounds were identified:\footnote{\textit{Ibid} at para 40.} that the police failed to exercise the discretion recognised to them by s 40(1)(j) of the CPA,\footnote{CPA s 40(1)(j) reads: ‘[a] police officer may without a warrant arrest any person who ... wilfully obstructs him in the execution of his duty’.} and that the arrest was contrary to s 28(2) of the Constitution. Exercise of discretion meant that the police officers had to consider and weigh the relevant circumstances to decide whether arrest was necessary and justified.\footnote{\textit{Raduvha} (note 4 above) at paras 42–43.} While they have the power to arrest, they are not obliged to do so. The exercise of police discretion is affected by the Bill of Rights,\footnote{\textit{Ibid} at para 47.} including s 28(2). According to the Court, even if the jurisdictional facts in s 40 of the CPA are satisfied, police need ‘to go further and not merely consider but accord the best interests of such a child paramount importance’.\footnote{\textit{Ibid} at para 48.} The police officers in this case were indifferent to the applicant being a child,\footnote{\textit{Ibid} at para 51.} not being a danger and being under parental care at the time of arrest.\footnote{\textit{Ibid} at para 52.} The Court found that this approach to the arrest of a child was incompatible with s 28(2) of the Constitution.\footnote{\textit{Ibid} at para 58.}

Section 39(2) of the Constitution requires that the interpretation of legislation be infused with the values promoted by the Constitution,\footnote{\textit{Ibid} at para 49.} and be informed by s 28(2) which ‘seeks to insulate them [children] from the trauma of an arrest by demanding in peremptory terms that, even when a child has to be arrested, his or her best interests must be accorded paramount importance’.\footnote{\textit{Ibid} at para 57.} In effect, the Court supported the view that the arrest of the child should be rarely done (although the Court fell short of declaring it a last resort): ‘an arrest of a child should be resorted to when the facts are such that there is no other less invasive way of securing the attendance of such a child before a court’.\footnote{\textit{Ibid} at para 59.} Although the best interests of the child had to be given paramount importance in relation to the arrest of a child, it would not at all times prevent such arrest. What the Constitution requires is a child-sensitive criminal system which does not react disproportionately to a child’s misbehaviour.\footnote{\textit{Ibid} at para 59.}
The Court rejected the amicus argument that s 28(2) of the Constitution should be made an additional jurisdictional requirement to those in s 40 of the CPA. It decided that s 28(2) can be given effect in the interpretation of s 40(1) of the CPA, as required by s 39(2) of the Constitution.\textsuperscript{132}

Section 28(2) was interestingly used in \textit{Raduvha}, as the only Bill of Rights provision relevant for assessing the lawfulness of a child’s \textit{arrest}. The Court proposed both an abstract use of s 28(2) (in the interpretation of a statutory provision which contained no special rules in relation to the arrest of a child) and a concrete use (in assessing the lawfulness of the exercise of police discretion in relation to the arrest of a specific child). It does not seem that the Court relied on the right provided in s 28(2), although it clearly applied that section independently of any other constitutional standard. Indicative of the fact that s 28(2) was not approached as a right is the fact that the Court did not refer to \textit{Fitzpatrick}, or to s 36 of the Constitution.

Although no substantial clarification is offered in \textit{Raduvha} in relation to s 28(2) as a right, the case advances the jurisprudence on the independent normative value of this section. Section 28(2) was used to secure a child-sensitive interpretation of the law and to direct the exercise of discretion on a child-sensitive path where no other, more explicit constitutional standard was applicable. This reliance on s 28(2) as a fall-back provision is apparent when the reasoning of the Court in relation to the \textit{arrest} of the child is compared with that concerning the \textit{detention} of the child. In the latter case, the Court simply applied s 28(1)(g) whose wording covered child detention.

C The significance of the two cases for the jurisprudence on the best interests of the child

In the two cases, the Court approached s 28(2) from two different perspectives: in \textit{J v NDPP}, as an independent right; in \textit{Raduvha}, as a principle or constitutional standard which informed the interpretation of a statute and the way police must exercise the discretion to arrest a child. In both cases the Court’s decision was driven by s 28(2). It follows that this section can be applied independently of other constitutional provisions regardless of the function it plays: as an interpretation tool/principle of interpretation or as an independent right. In both cases and in both capacities, the application of s 28(2) led to a child-favourable outcome: the invalidation of a statutory provision or a child-focused interpretation of a statutory provision which did not consider the special position of children in conflict with criminal law.

In both cases s 28(2) seemed to have been the only or the most relevant constitutional provision, which means that the Court had little choice in terms of the constitutional provisions it could apply. Does this mean that s 28(2) is to be applied independently only in subsidiary situations,\textsuperscript{133} when other, more specific s 28(1) sub-sections are not relevant? In the past, Skweyiya J hinted at this approach in \textit{Le Roux}, where, to justify reliance on s 28(2), he mentioned that ‘none of the rights listed in section 28(1) have direct bearing here’.\textsuperscript{134} This subsidiary approach to s 28(2) may also be supported by Kampepe J’s judgment in \textit{Teddy Bear}. There, the Court applied ss 10 and 14 of the Constitution, and, as a third leg of its reasoning, assessed the constitutionality of the impugned statutory provision against s 28(2).

\textsuperscript{132} Ibid at paras 63–65.

\textsuperscript{133} It is not argued here that the consideration of the best interests of the child is of a subsidiary nature. The subsidiarity as discussed here refers only to the independent application of s 28(2) as a right.

\textsuperscript{134} \textit{Le Roux} (note 94 above) at para 210.
As mentioned in part IV, none of the rights in s 28(1) was relevant in that case, and thus the Court found a violation of s 28(2) in the teenagers’ exposure to the harm created by coming into contact with the criminal justice system. The Court has not made a statement of principle in relation to a potential subsidiarity approach to the application of the best interests, and it remains to be seen if this is what the Court intended with these cases.

Following this approach would introduce some much-needed certainty in relation to the application of s 28(2). However, judicial statements contrary to a subsidiarity approach can also be found in the case-law. For example, Skweyiya ADCJ said in the J v NDPP that ‘the starting point for matters concerning the child is section 28(2)’, suggesting that, rather than being a subsidiary legal ground, s 28(2) should be the ‘starting point’ in a legal enquiry. This may explain the preference of the Court for s 28(2) as a legal ground for its judgment in J v NDPP, despite other constitutional provisions having been invoked in that case. The Court did not fully explain its reasons for doing so. It may be that the Court simply found it easier to engage with s 28(2) given this section’s support in the Court’s jurisprudence and legal culture more generally. It may also be that, in a case in which the Court had to be mindful of differences between children and adult sex offenders, s 28(2) provided solid grounds to deliver a child-focused judgment and to insulate its reasoning from subsequent challenges to the legislation by adult offenders.

Like many aspects concerning the best interests of the child, the ‘starting point’ expression used by Skweyiya ADCJ invites further reflection. J v NDPP is not the first case in which this Justice used the ‘starting point’ phrase, although in this case it received the approval of the majority. In Le Roux, Skweyiya J said (after finding that no right in s 28(1) was relevant) that s 28(2) –

forms the basis and starting point from which the matter is to be considered. Once the considerations relevant to this foundation are clearly cemented, one can then begin to examine the other rights that enter the balance, without losing sight of the fact that the best interests of the child remain ‘of paramount importance’.

The question which immediately arises from Skweyiya J’s position is whether the ‘starting point’ approach applies only to matters directly concerning children (as in J v NDPP or Le Roux) or also to matters concerning children only indirectly but affecting them (as in M v S or Pridwin Preparatory School).

This may be an important issue to address, and one that could further unpack the application of s 28(2). Following Skweyiya J in J v NDPP, it can be argued that the best interests of the child should be the starting point in matters concerning a child directly, if no other more specific constitutional provision is at stake. In that case, the independent right function of s 28(2) can be relied on. Conversely, in a matter indirectly concerning the child, the best interests of the child should not be the starting point. The importance of establishing when the best interests of the child are paramount as a starting point, and when they are not, is apparent in the limitation inquiry. Thus, if s 28(2) is the starting point, this section ought to be given attention as a right (its content must be established by the courts and any limitation of the right must comply with s 36). This is the highest normative position that s 28(2) can occupy.

135 J v NDPP (note 3 above) at para 35.
136 Similar point in Skelton (note 18 above).
137 Le Roux (note 94 above) at para 210.
138 Ibid at para 211.
When s 28(2) is not the starting point, it will operate as a justification to limit other rights. Then, courts may not need to give structured attention (i.e. give full clarity on legal obligations and entitlements) to the right in s 28(2) because s 36 allows for a limitation of rights on other than rights-based grounds (i.e. the limiting factor need not always be a right). Thus, s 28(2) can justify a limitation of rights under s 36 by operating as a principle or as a standard.

The ‘starting point’ approach of Skweyiya J has clear benefits for children as it directs the judges to conceptualise relevant cases starting from the rights of children, regardless of the existence of other competing interests. This means that in a case concerning a child, a court will establish the legal issues to be decided with reference to the best interests of the child/rights of children, after which the other interests concerned will be considered, possibly to establish whether they can limit the best interests or the rights of the child. By its nature, such an approach makes the rights of children the focal point of the judicial inquiry. The majority judgment in Centre for Child Law v Media 24 is an example of a case where the starting point approach was not followed, and this is reflected in how the Court dealt with the matter. The ‘starting point’ for the majority was freedom of expression, to which the Court gave substantial attention,139 and the interests of children were considered only as factors able to justify a limitation of this freedom.

The ‘starting point’ approach was mooted by Skweyiya J in Le Roux and J v NDPP, which were both matters concerning children directly. The question arises as to whether it can (or should) be extended to matters concerning children indirectly. If I am correct in my understanding of the ‘starting point’ approach, the difficulty turns to establishing what concerns children directly and what concerns them indirectly. Cases like Le Roux and J v NDPP concerned children directly, but cases like M v S and Pridwin Preparatory School concerned them indirectly.140 However, the Court has not yet clarified this aspect. In M v S, Sachs J acknowledged the risk of s 28(2) being ‘spread too thin’141 and thus losing its protective function, but did not offer a solution to minimise this risk, leaving the scope of this section to be decided on a case-by-case basis.

Another question raised by the ‘starting point’ approach is whether it pertains only to cases in which s 28(2) applies as an independent right (as was the case with J v NDPP and Le Roux per Skweyiya J); or whether it applies when s 28(2) is not applied independently, but is associated with other rights of children. Confining the application of this approach only to when s 28(2) is applied independently as a right creates a hierarchy between the rights in s 28. This is undesirable and unlikely to have been envisaged by the drafters of the Constitution. At the same time, if the starting point approach is used in all cases concerning children (directly and indirectly; or when s 28(2) applies by itself and when it applies in association with other rights), this may result in an artificial elevation of the rights of children above all other rights or other legitimate interests. Such elevation has been consistently rejected by the Court.142

It was mentioned above that one of the difficulties with the Court’s best interests jurisprudence is that it is not easy to ascertain how it uses s 28(2) (as a right, a principle, a standard, or an interpretive tool). J v NDPP and Raduwa may help with distinguishing

139 Centre for Child Law v Media 24 (note 95 above) at paras 14 and 16.
140 Pridwin Preparatory School (note 92 above) (the starting point was the best interest of the child only because the parties agreed that it is so).
141 M v S (note 7 above) at para 25.
142 The Court noted that children’s rights can be limited in terms of s 36 and that an over-extension of s 28(2) may undermine its effectiveness. Ibid.
between these different uses. In *J v NDPP* the Court engaged with s 28(2) on the basis that it contained an independent right. The ‘markers’ of such use were the identification of legal requirements arising from that section and the use of s 36 to test the legitimacy of potential limitations to it. The latter is especially significant because a section-36 inquiry is employed only when limiting rights in the Bill of Rights and not when, for example, certain limitations are considered on the application of legal principles.

In *Raduvha*, the Court applied s 28(2) independently but not as a right. In *Raduvha*, the Court did not refer to precedents in which it declared that s 28(2) contains an independent right and did not reiterate its position that it can be limited as per s 36. Two types of usage of s 28(2) can be distinguished in *Raduvha*: an abstract use (which led to a Constitution-compliant interpretation of the relevant provision in the CPA) and a concrete use (the Court assessed how the arresting officers gave effect to the best interests of Ms Raduvha when deciding her arrest). *Raduvha* illustrates the coexistence within s 28(2) of the new, human rights capacity of the best interests of the child standard (manifested in this case in the control exercised over the meaning of a statute by directing a constitutional interpretation) and of the traditional, common law-like capacity of the standard, that of a standard enabling a decision-maker to make decisions concerning a specific child/children taking into consideration all the relevant factors.

VI CONCLUSIONS AND SUGGESTIONS FOR THE WAY FORWARD

*J v NDPP* and *Raduvha* take small steps toward the clarification of the best interests of the child jurisprudence of the Court. Full clarity can only develop gradually given the multitude of roles played by the best interests of the child, the variety of legal sources operating with the concept, and the currently unsystematic jurisprudence dealing with it. The two cases illustrate the coexistence and the complementarity of the functions played by the best interests concept. They arguably support a subsidiary application of s 28(2) both as an independent right and as a constitutional principle when no other s 28(1) rights are applicable. Apart from making the application of s 28(2) more predictable, this approach would also safeguard against the normative bloating of this section, to the detriment of the rights in s 28(1).

In *Fitzpatrick* and subsequent cases, the Court expressed its reluctance to define the content of s 28(2) to preserve its flexibility to respond to the individual circumstances of each child. It seems, however, contrary to the very essence of rights (as legal tools) not to define them and thus rob them of the potential advantages which result from this status. In the context of the best interests of the child, its recognition as a right represents a chance to rehabilitate a concept which has been often misused to the advantage of adults. Refusing to define the legal content of s 28(2) or of the right identified therein creates the risk of recycling the paternalistic version of the best interests with its ills.

Arguably, the reluctance to define the content of s 28(2) is due to an insufficient unpacking of the concept, both judicially and academically. There are many issues which require further analysis. Without intending to be exhaustive, a few can be suggested: the relationship between the constitutional, statutory and common law concepts of the ‘best interests of the child’, including overlaps, distinctions, similarities and reciprocal influences; distinctions, similarities and scope of application of s 28(2) as a right in itself and as a principle/guideline/standard; and the meaning of ‘flexibility’ in the context of the best interests of the child. A few thoughts
are presented below on the issue of flexibility to illustrate the need to unpack terms associated with the best interests of the child and perhaps too easily assumed as self-evident.

The complexity of the concept of the best interests of the child means that flexibility can relate to many different issues such as the meaning of the concept through to the outcome of its application and factors influencing its determination. The meaning of ‘flexibility’ itself ought to adapt to the changed nature of the concept. Thus, the flexibility of the best interests as a constitutional/human rights concept should be distinguished from the flexibility of the concept as understood at common law. In this way, the inherent flexibility recognised in the concept at common law (in relation to, for example, a non-exhaustive list of factors influencing it and its ability to secure the best possible outcome for an individual child) can co-exist with a different kind of flexibility which is specific to the constitutional/human rights dimension of the concept. Arguably, this nuanced approach to the ‘flexibility’ of the best interests of the child rests on acknowledging that this concept has both factual and juridical dimensions.

Some judicial support can be identified for this differentiated approach. In *M v S*, Sachs J said that ‘the indeterminacy of outcome is not a weakness\(^{143}\) of the best interests of the child standard. The ‘indeterminacy’ of the best interests was thus associated with its outcome in specific cases and not the legal content of s 28(2). Thus, while the best interests in specific matters cannot be prescribed or pre-determined through general/abstract statements (in that sense, the best interests of a child is flexible) this is not incompatible with knowing what obligations and entitlements arise from s 28(2) (as a constitutional provision with an ascertainable legal content). In the latter sense, the flexibility of the constitutional best interests of the child cannot be much different to that inherent in other human rights norms, whose content develops through judicial application.

In *J v NDPP* the Court broke with its reluctance to spell out the content of the right of children to have their best interests given paramount importance, identifying some specific requirements arising from s 28(2). It demonstrated that doing so does not undermine the capacity of this section to respond to individual circumstances and does not prevent it from securing child-focused legal outcomes. Giving contour to the legal content of s 28(2) is not therefore inimical to its flexibility, if the requirements do not prescribe *in abstracto* a certain outcome for individual children and do not prevent the further development of its content. Continuing to avoid establishing the content of s 28(2) suggests that this concept is not epistemologically legal (i.e. cannot be known and understood through legal means). This is, arguably, contrary to the rule of law, and also an unnecessary path, as the Court’s own jurisprudence illustrates that s 28(2) is amenable to a better-defined legal content. Arguably, acknowledging the different functions or dimensions of the best interests of the child, from which different normative requirements arise, may be a good start. It is useful for the Court to acknowledge that the best interests of the child test is an umbrella term, which has by now accumulated an expansive meaning which requires some systematising to improve its functionality.

The two cases illustrate the coexistence and the complementarity of the functions played by the best interests. They arguably support a subsidiary application of s 28(2) both as an independent right and a constitutional principle when no other s 28(1) rights are applicable. Apart from making the application of s 28(2) more predictable, this approach would also safeguard against this section engulfing the other rights in s 28(1).

\(^{143}\) *M v S* (note 7 above) at para 24.
As mentioned before, s 28(2) (however applied) has had a significant positive impact on South African law. A challenge arises: how can the law be taken forward and clarified without reversing the gains made with the application of s 28(2) so far? Central to the current achievements of this section has been its ability to secure a child-specific legal treatment, sensitive to the needs of children. This is significant and it is important that all children benefit from such treatment and that s 28(2) therefore be applied consistently. For this to occur, the law has to develop in such a way that judges see the application of s 28(2) as a matter of legal obligation, and not one of philosophical disposition.
Kindred Strangers: Why has the Constitutional Court of South Africa Never Cited the African Court on Human and Peoples’ Rights?

TOM GERALD DALY

ABSTRACT: Why has the Constitutional Court of South Africa never cited the African Court on Human and Peoples’ Rights? The two courts appear to be natural allies, having both elaborated a robust jurisprudence promoting civil-political and socio-economic rights, accountability, political participation, and good governance. However, despite the African Court having issued a raft of landmark merits judgments since June 2013, the Constitutional Court has yet to cite its jurisprudence. This article attempts to account for this apparent lacuna in South African case-law, placing it against the Constitutional Court’s overall approach to citing international law and courts, and suggesting a range of possible explanatory factors, including: the state’s position as a ‘reluctant regionalist’; institutional factors (such as the Constitutional Court’s possible preference to retain constitutional supremacy and adjudicative autonomy, and tendency to more readily cite non-African jurisprudence); and broader structural factors (such as a lack of citations in submissions to the Court). It is argued that this matters for two reasons. First, it may possibly deprive the Constitutional Court of sources that could enrich its jurisprudence and anchor it in the developing regional human rights system. Second, as perhaps the two most important courts on the entire continent as regards rights protection, it seems desirable that the relationship between the Constitutional Court and the African Court should be developed and deepened – which does not mean they will always agree.

KEYWORDS: Constitutional Court of South Africa; African Court on Human and Peoples’ Rights; constitutional law; international human rights; good governance; judicial dialogue

AUTHOR: Assistant Director, Melbourne School of Government. Email: thomas.daly@unimelb.edu.au

ACKNOWLEDGEMENTS: Sincere thanks to Stu Woolman for the invitation to write this article, and to Justice Albie Sachs, Micha Wiebusch, and Michael Bishop for their valuable insights, as well as other participants at the Constitutional Court Review IX conference, and the two anonymous reviewers. I am also grateful to judges and staff at the African Court of Human and Peoples’ Rights, as a portion of research for this article was conducted as part of a consultancy report to design an African Judicial Network in late 2017.
I INTRODUCTION: A STRANGE PROPOSITION?

At first blush, it may seem a strange proposition to enquire about the relationship between the Constitutional Court of South Africa and the African Court on Human and Peoples’ Rights (‘the African Court’). After all, South Africa as a state has limited engagement with the African Court. Despite having ratified the Protocol establishing the Court in 2002, the state has not made the special declaration required to permit individuals and non-governmental organisations (NGOs) to petition the Court. The African Court has not yet issued any judgment in a case against South Africa.

Those issues, however, relate solely to the formal relationship between the state as a whole and the African Court, and are not the main focus here. What drives the enquiry in this article is to explore why the South African Constitutional Court itself has not fostered a particularly strong relationship with the African Court, and what this tells us about the self-perception of the South African Court and its perception of, or attitude toward, international judicial power within its own region, as well as the structural barriers to a closer relationship. The weakness of the relationship is evidenced, in particular, by the fact that the South African Court has not cited the African Court’s jurisprudence even once in the almost six years since the latter’s first merits judgment in June 2013.\(^1\) It is also reflected in the literature on the South African Court: even the most recent works on the Court make no mention of the African Court.\(^2\) In fact, no account of the Constitutional Court’s citation of the African Court yet exists – nor any full account of the wider relationship between the Constitutional Court and the African Court, or the Constitutional Court’s relationship more broadly with the wider African human rights system.\(^3\) This represents a significant gap in the literature.

This article attempts to account for this apparent lacuna in South African case-law, placing it against the Constitutional Court’s overall approach to citing international law and courts, and arguing that it cannot be simply explained by structural factors, including the fact that South Africa has yet to make the special declaration required to permit individual and NGO petitions to the African Court (which greatly limits the state’s interaction with the Court), or that the African Court has not issued any judgment regarding South Africa. Rather, a range of other possible explanatory factors appear to be at play, including: the state’s position as a ‘reluctant regionalist’; institutional factors (primarily, the Constitutional Court’s possible preference to retain constitutional supremacy and adjudicative autonomy, tendency to more readily cite non-African jurisprudence, and the African Court’s youth); and broader structural factors (such as a lack of citations in submissions to the Court and a civil society view of the African Court as an alien entity).

The article’s main claim is that this matters for two reasons. First, it may possibly deprive South African jurisprudence of sources that could enrich it and anchor it in the developing regional human rights system. Second, as perhaps the two most important courts on the entire continent as regards rights protection, it seems desirable that the relationship between the Constitutional Court and the African Court should be developed and deepened. The South

---

1 This lacuna in the jurisprudence is discussed below in part IV.
African Constitutional Court has a uniquely influential position in the African Union (AU), as regards other apex national courts, while the African Court has growing importance as the final judicial interpreter of the African Charter of Human and Peoples’ Rights – which, as the most widely ratified rights treaty in the AU (54 of the 55 AU member States to date, except the Kingdom of Morocco, which re-joined the AU in January 2017), may arguably be viewed as the primary pan-continental human rights instrument.4

The article contains four substantive parts. Part II briefly sets out the methodology followed and necessary caveats. Part III examines the development and purposes of the South African Constitutional Court and the African Court, highlighting the commonalities and divergences between the two institutions, and arguing that the two are ‘natural allies’ with similar approaches. Part IV examines the South African Constitutional Court’s existing approach to the citation of international law and jurisprudence, highlighting that its openness to international norms is partial and skewed towards certain courts, and has made relatively little room for African human rights instruments and jurisprudence. Finally, part V canvasses the possible reasons why the South African Constitutional Court has not developed a strong relationship with the African Court.

II METHODOLOGY AND NECESSARY CAVEATS

In order to assess whether, and how, the Constitutional Court cites African Court jurisprudence, and the African Charter more generally, this article has utilised a relatively simple methodology. Every judgment of the Constitutional Court since 14 June 2013 (the date of the African Court’s first merits judgment) until the time of writing (25 February 2019) has been searched on the South African Legal Information Institute (SAFLII) database5 using four search terms: ‘African Court’, ‘African Charter’, ‘charter’, and ‘Banjul’ (the latter to catch any reference to the African Charter as the ‘Banjul Charter’).6 The initial aim was to simply identify how many times the Constitutional Court had cited the case-law of the African Court, and then analyse these instances in more depth. Finding that the Constitutional Court had not yet cited any African Court judgment was unexpected, and sent the analysis in a different direction.

The broader methodological thrust of this enquiry is comparative, in the sense that it is couched in an understanding of experiences in other world regions with continental human rights courts – namely, Latin America and Europe – where, despite dominant positive accounts of regional judicial interaction,7 national courts have often difficult relationships with the regional international human rights court.8 These inter-court relationships run the gamut from highly engaged, to patchy, to even hostile or acutely undeveloped. The approach in this article might also be called ‘thinly’ normative, in the sense that it suggests that the Constitutional Court should, at least, pay greater attention to the case-law coming from the African Court.

---

4 The African Charter on the Rights and Welfare of the Child has been ratified by 48 state parties. However, its breadth does not quite compare with that of the Banjul Charter.
6 At the time of writing, the most recent judgment of the Constitutional Court on SAFLII was Buffalo City Metropolitan Municipality v Metgosiv (Pty) Limited [2019] ZACC 9.
8 For a discussion of this common tension between domestic courts and international tribunals, see part V below.
in Arusha given its position as the final judicial interpreter of the African Charter on Human and Peoples’ Rights as well as a range of other international instruments.

A number of caveats are required here. First, there is no attempt to crudely analyse the African context through Inter-American and European lenses. The African context must be analysed on its own merits. Nor is there any form of inchoate teleology underlying this analysis, suggesting that the African human rights multi-level judicial system does – or ought to – develop in a manner similar to either the Inter-American and European systems. Each of those systems evidently has its own strengths and weaknesses, based on the wider historical, socio-political and regional context in which it has developed. Nor should the analysis be understood as suggesting a unidirectional flow of norms from the international to the domestic level: the exchange is not only bidirectional, but tends to be multidirectional, with national courts influencing human rights courts (eg, Colombian and US Supreme Court case-law influencing the Inter-American Court), and some national courts (eg, the constitutional courts of South Africa, Colombia, and Germany) influencing how apex courts in other states approach, and give effect to, international human rights jurisprudence. There is also no argument here that the Constitutional Court should slavishly or uncritically follow African Court jurisprudence or ‘shoehorn’ the latter’s case-law into analysis of the domestic Constitution. Nor can there be any facile analysis of one court’s jurisprudence necessarily being superior to the other. These courts have different missions, audiences and resources, operate in different contexts, and under different political pressures and constraints.

That said, it is worthwhile to acknowledge that the African Court has itself made repeated references to both the Inter-American and European human rights systems and courts in its jurisprudence and has pursued inter-regional judicial collaboration in various ways beyond its case-law, especially through organising face-to-face meetings with judges from each system. It also may be argued that, while remaining cognisant of the differences between the African continental human rights system and the Inter-American and European systems, the variety of relationships between national courts and human rights courts in these other regions is illuminating in itself, emphasising that there is no one model for ideal interaction.

Analysis in other regions, by scholars such as Alexandra Huneeus (Latin America) and Nico Krisch (Europe) highlights that many different factors come into play in shaping how national courts interact with, and perceive, international human rights courts. Krisch highlights, in particular, that broadly positive interaction in Europe is not achieved through any formal design, but by each court engaging in ‘judicial diplomacy’ and strategy across the national/international divide to allow the overall system to function without excessive

---


11 This proposition has held true since the Court’s first judgment. Mtikila v Tanzania ACHPR, App. 009/2011 and 011/2011 (14 June 2013).

friction. Huneeus suggests that these relationships are strongly conditioned by judicial culture and domestic courts’ self-perception of the ambit of their role, judicial attitudes to regional (judicial) oversight and a wish to retain the ‘final say’ in constitutional matters, executive influence (eg, in Venezuela), the overall ideological leanings of the state, the formal status of international law under the constitution, and even familiarity with regional jurisprudence and international law. While in no way determinative, these insights are useful in discussing the relationship between the Constitutional Court and the African Court in this article (bearing in mind also that the IACtHR operates alongside a regional commission on human rights while the ECtHR does not).

Finally, it may be argued that this enquiry should focus equally on the African Commission on Human and Peoples’ Rights: the Commission has been in operation since 1987 – seven years before the South African Constitutional Court was established – and it has co-equal footing with the African Court. However, the aim here is to focus more squarely on the distinct relationship between the Constitutional Court and the African Court as judicial institutions. The Commission, for all its powers and centrality within the African Union’s human rights system, remains a quasi-judicial institution that is neither formally, nor in full practice, a court. Needless to say, the enquiry here could usefully be expanded to discussion of the Commission in further research. This article also forms the basis for potentially illuminating comparison between the Constitutional Court’s record of citing African Court jurisprudence and that of comparable courts, such as the Constitutional Court of Benin or the Supreme Court of Kenya, which cannot be pursued here.

III KINDRED SPIRITS: COMPARING THE SOUTH AFRICAN CONSTITUTIONAL COURT AND THE AFRICAN COURT

In many ways, the design and development of the South African Constitutional Court and the African Court mark them out as kindred spirits. Both are charged with a similar (though not identical) function to interpret a central document strongly focused on human rights and good governance, and both have energetically seized their mandates to elaborate a jurisprudence that does not shy away from taking assertive stances and speaking truth to power. However, the Courts are of different vintages: the South African Constitutional Court is more than 20 years older than its regional peer. This places them in different positions of recognition and power, and has clear implications for their relationship. This section briefly provides an account of each Court’s trajectory to date, and its position within its own institutional setting, as well as its regional impact.

A The South African Constitutional Court

From the very beginnings of South Africa’s transition to democratic rule after minority governance under apartheid in the early 1990s, the Constitutional Court has been a central institution. The final Constitution, produced by a Constituent Assembly after the 1994 elections and which entered into force in December 1996, enshrined a number of countermajoritarian mechanisms aimed at placing constitutional fetters on the African National Congress (ANC)

and providing guardrails for the fledgling democratic order. The constitutional text expressly states the political system to be based on the values of human dignity, equality, human rights, the supremacy of the Constitution and the rule of law, and a ‘multi-party system of democratic government, to ensure accountability, responsiveness and openness’.15

Much ink has been spilled on the placement of Constitutional Court as a central actor in the democratic constitutional order, with a wide range of powers aimed at constraining political powers, guarding the separation of powers, and upholding a long raft of fundamental rights. As a constitutional design option, the Court was designed to act as a constraint on the electoral dominance of the ANC and as protector of the white minority’s rights in the new black-majority political system, and, as such, constituted a central guarantee in the political settlement underlying the democratic transition and the new constitutional order – as well as indicating concrete commitment to the grand ideals in new democratic constitution for a more just and equal society.16 As Klug has observed, the Court ‘has been called upon to address issues and to face challenges that would be considered extraordinary for any judiciary.’17

The Court quickly cemented its reputation for assertiveness in the 1990s with decisions holding the death penalty to be unconstitutional, ordering the enactment of laws on same-sex marriage in line with the Constitution, and upholding prisoners’ voting rights.18 The Court also developed an innovative, and internationally recognised, jurisprudence aimed at striking an extremely difficult balance between attempting to deliver on the promises of democracy and social justice in the 1996 Constitution, and avoiding overstepping the bounds of possible (and democratically proper) action in South Africa’s democratic system – although it has been criticised for taking a less robust approach to upholding social and economic rights than other courts, such as the Colombian Constitutional Court.19

The Court has repeatedly insisted that human dignity, equality, freedom, and individual rights, repeatedly proclaimed within the text, are to be viewed not as subtracting from the democratic principle, but rather, lying in ‘constructive tension’ with majority rule.20 The Court has also indicated its rejection of any winner-takes-all conception of majority rule and has emphasised the need for a deliberative democracy where the minority as well as the majority are included in public decision-making.21 The successes of the Constitutional Court in constraining the government, and the government’s apparent willingness to abide by the Court’s rulings, have been central to the perception of a positive trajectory in the crafting of a functioning democratic order underpinned by a robust rule of law.22

15 Constitution s 1.
18 Roux (note 16 above) at 235–364.
21 Ibid.
22 Fowkes (note 2 above).
More recent judgments in the Democratic Alliance,23 Glenister,24 and Nkandla25 cases, pushing back against perceived attacks on the Constitution by the Zuma administration – focused on anti-corruption agencies and presidential corruption in particular – have cemented the Court’s reputation as a defender of the constitutional system and a key guarantor of the separation of powers.26 Internationally, as Law and Chang have noted, the Constitutional Court is one of the few apex courts of the Global South that have entered the pantheon of globally-recognised and cited courts, alongside other ‘premier’ courts such as the US Supreme Court, Canadian Supreme Court, and the Federal Constitutional Court of Germany.27 The South African Constitutional Court is also a judicial leader in its own region, having a significant influence on the jurisprudence of other courts, such as those of Kenya and Uganda,28 and also serving as an institutional model for newer apex courts, such as the constitutional courts of Zambia and Zimbabwe.29

B The African Court on Human and Peoples’ Rights

Following a long period of advocacy by academics and NGOs the AU adopted a protocol in 1998 to establish an African Court on Human and Peoples’ Rights.30 The protocol did not enter into force until January 2005, and the Court was finally established on 2 July 2006, as its first eleven judges were sworn in before a summit meeting of African leaders in the Gambian capital, Banjul.31 It is based in Arusha, in northern Tanzania.

To date, 30 states have recognised the Court’s jurisdiction – more than half of the AU’s 55 member states.32 However, to date a mere nine states have made the requisite declaration to allow direct individual and NGO petitions to the Court,33 and one, Rwanda, rescinded this recognition in early 2016.34 This slow uptake has significantly limited the scope of the Court’s material jurisdiction, and is one factor in the seven-year wait for its first merits decision in a

25 Economic Freedom Fighters v Speaker of the National Assembly and Others; Democratic Alliance v Speaker of the National Assembly and Others [2016] ZACC 11, 2016 (3) SA 580 (CC).
27 As Law & Chang show, the Court is still cited far less by other apex courts that those it chooses to cite frequently.
contentious case.35 The Court has generally been reliant on the referral of cases by the African Commission on Human and Peoples’ Rights, which appeared to evince considerable reluctance in the early years – referring only two cases before 2012.36 That the African Court has been somewhat overlooked in the South African context is therefore, in one way, unsurprising.

However, the African Court has quickly developed a robust, high-quality and assertive jurisprudence in the almost six years since its first merits judgment. This summary focuses on the 23 merits judgments issued by the Court at the time of writing, which can be divided into four broad themes:37

(i) Political Participation: In its first merits judgment, issued in June 2013 in Mtikila v Tanzania,38 the Court unanimously found the ban on independent electoral candidacies in Tanzania’s national Constitution to constitute a violation of the African Charter. In late 2016 the Court ruled in APDH v Côte d’Ivoire39 that a new law on the Electoral Commission violated both the right to equal protection of the law in art 3(2) of the African Charter on Human and Peoples’ Rights and art 10(3) of the African Charter on Democracy, Elections and Governance for placing opposition electoral candidates at a disadvantage by packing the body with representatives of the President, government ministers and the President of the National Assembly (Parliament).

(ii) Freedom of Expression: In March 2014, in Zongo v Burkina Faso40 the Court found the state in violation of rights to judicial protection and free speech for failing to investigate and prosecute the killers of a journalist and his companions in 1998. In December 2014, in Konaté v. Burkina Faso41 the Court unanimously ruled a twelve-month sentence of imprisonment for criminal defamation imposed on the applicant journalist in 2012 (for having accused a public prosecutor of corruption) to be a violation of the Charter right to freedom of expression. In November 2017 in Ingabire v Rwanda,42 the Court deemed Rwanda in violation of the free speech rights in the African Charter (art 9(2)) and the ICCPR (art 19) and rights to an adequate defence under art 7 of the African Charter due to a 15-year sentence of imprisonment imposed on the applicant, an opposition leader, for crimes including spreading genocide ideology, complicity in acts of terrorism, sectarianism, and terrorism in order to undermine the authority of the State.

(iii) Fair trial, Liberty, and Equal Protection before the Law: In the Thomas,43 Onyango44 and Abubakari45 cases against Tanzania, decided in 2015 and 2016, the Court found the state in violation of the right to a fair trial in art 7 of the African Charter in each case. In the Saif Al-Islam Gaddafi46 judgment of June 2016 – the first referred by the African Commission – the Court found

35 To date, only Benin, Burkina Faso, Côte d’Ivoire, Ghana, Malawi, Mali, Rwanda and Tanzania have made the required declaration. Rwanda has since withdrawn its declaration.
37 Other judgments that do not clearly fit within these four themes are Koura and Diabaté v Mali, ACHPR, App. Nop. 040/2016 (21 March 2018)(classification of a criminal offence) and Anudo v Tanzania, ACHPR, App. No. 012/2015 (22 March 2018)(deportation and right to citizenship).
41 ACHPR, App. No. 004/2013 (5 December 2014).
45 ACHPR, App. No. 007/2013 (3 June 2016).
the secret detention and criminal proceedings against the second son of former Libyan President Gaddafi in violation of articles 6 (right to personal liberty, security and protection from arbitrary arrest) and 7 (right to fair trial). In late 2017 the Court issued three further merits decisions. In Jonas v Tanzania and Onyachi v Tanzania the Court again found the State in violation of the rights to, respectively, fair trial (art 7 of the African Charter) and liberty (art 6). Adding to its previous judgments in the Thomas, Abubakari and Onyango cases, the Court’s case-law has developed a pattern of sustained criticism of the deficiencies Tanzania’s criminal justice system, concerning free legal aid, timely issuance of trial judgements, organisation of identification parades, and appropriate consideration of defences forwarded by the defendant.48 However, it has not found a violation in every case: in Isiaga v Tanzania,49 for instance, it found no violation where the applicant alleged his fair trial right had been breached due to erroneous visual identification and rights against discrimination arising from the refusal of legal aid; and in Viking and Nguza v Tanzania,50 it found no violation due to a lack of sufficient evidence of bias and collusion in the applicant’s trial for sexual offences. The Court’s focus on fair trial and due process issues, especially in Tanzania (eg, free legal aid, retrospective legislation), now forms a routine and central part of its jurisprudence, as seen in more recent decisions.51

(iv) Social and Economic Rights of Indigenous Communities: In the landmark Ogiek case against Kenya in May 2017 – referred to the Court by the Commission on the basis that it concerned serious and massive rights violations – the Court held that the Kenyan government had violated no less than seven articles of the African Charter, including collective rights, in a far-reaching dispute concerning the ancestral lands of the Ogiek community. Building on, and largely agreeing with, previous African Commission decisions in similar cases, the Court found violations of the rights to non-discrimination (art 2), culture (art 17(2) and (3)), religion (art 8), property (art 14), natural resources (Article 21) and development (art 22). The judgment has been interpreted as recognising, in practical terms, a right to land, a right to food, and, potentially, a right to free prior and informed consent regarding state interference with ancestral lands.53

The Court’s jurisprudence is notable for a range of reasons. In particular, compared to the comparatively more quiescent (albeit far from universally quiescent) African Commission on Human and Peoples’ Rights, the Court’s jurisprudence shows a marked willingness to

51 William v Tanzania, ACHPR, App. No. 006/2015 (21 September 2018)(Finding a violation of art 7 of the Charter for failure to provide free legal aid or hear defence witnesses, and convicting the applicant on the basis of insufficient evidence); Paulo v Tanzania ACHPR App. No 020/2016 (21 September 2018)(Finding no violation); Evarist v Tanzania, ACHPR App. No. 027/2015 (21 September 2018)(Finding an art 7 violation for failure to provide free legal aid); Makungu v Tanzania, ACHPR, App. No. 006/2016 (7 December 2018) (Finding an art 7 violation for failure, for over two decades, to provide the Applicant with copies of the records of proceedings and judgments in his criminal trial); Werema v Tanzania, ACHPR, App. No. 024/2015 (7 December 2018)(Finding no art 7 violation); Guehi v Tanzani, ACHPR, App. No. 001/2015 (7 December 2018)(Finding an art 7 violation for unreasonable delay in criminal proceedings and violation of art 5 of the Charter (right to dignity) for deprivation of food).
vindicate the rights in the African Charter and other instruments, particularly evidenced by the sheer number of violations found – and, in Mtikila, the Court’s willingness to find even a provision of a national Constitution incompatible with the Charter. The Court’s jurisprudence is also valuable for its broad comparative approach (drawing in particular on – but not slavishly following – the case-law of the Inter-American and European human rights courts, and the Human Rights Committee), as well as the way in which it has mitigated many of the starker deficiencies of the African Charter (compared to the American and European rights conventions). Most notably, in its first merits judgment in Mtikila, the Court reduced the impact of so-called ‘clawback clauses’ in the Charter through recourse to proportionality analysis – effectively establishing a ‘restriction on restrictions’. The Court has also clearly stated its power to order investigations and damages where necessary.

In line with its ability to interpret any rights treaty ratified by a respondent state, the Court has interpreted treaties such as the International Covenant on Civil and Political Rights (ICCPR) and the African Charter on Democracy, Elections, and Good Governance, as well as recognising the latter as a justiciable human rights instrument, which has bolstered its capacity to deal with sensitive electoral and governance issues in respondent states. The Court has also, at times, read different rights instruments jointly. In Zongo, for instance, the Court read art 66(2)(c) of the Revised ECOWAS Treaty (on journalists’ rights in general) in conjunction with art 9 of the African Charter (on the right to freedom of expression). These interpretations have given the Charter ‘teeth’ and amplified its strength as a human rights instrument for the entire African Union.

The Court has met with clear successes, such as the agreement of the Burkinabé authorities to open an investigation in compliance with the Court’s order in Zongo, and the recent judgment of the Lesotho Constitutional Court striking down domestic criminal defamation laws in line with the African Court judgment in Konaté. Konaté was also cited in a February 2017 High Court of Kenya judgment holding the law on criminal defamation to be unconstitutional. All of these developments, taken together, mark the African Court’s jurisprudence out as increasingly important, and as worthy of attention. It must be noted here, of course, that the Court in its work has often been building on the decisions of the African Commission.

Adem Abebe suggests that the court has indicated a willingness to act as, effectively, a ‘constitutional court for Africa’, although he observes that this has clear potential to provoke political backlash. Certainly, the Court has faced serious and multi-dimensional resistance to its authority, not least widespread refusal by respondent states to implement its decisions, including its host state, Tanzania. In addition, instruments geared towards institutional reform have left the African Court in a state of institutional insecurity, unsure whether it will be

---

54 Mtikila v Tanzania (note 38 above) paras 103–107.3.
55 Ibid at paras 106.1–106.5. It should be noted that the Court largely adopted the Commission’s approach in this regard.
56 APDH v Côte d’Ivoire (note 39 above) at paras 51–65.
57 Zongo (note 40 above) at para 180 (South Africa is not a party to the ECOWAS Treaty).
radically transformed: the Malabo Protocol adopted in 2014, if ratified, would merge the Court with the AU’s (not yet established) Court of Justice to create an African Court of Justice and Human Rights, and would expand the new court’s remit to international criminal jurisdiction.61

IV A LIMITED OPENNESS: THE CONSTITUTIONAL COURT’S APPROACH TO INTERNATIONAL LAW

This section explores the Constitutional Court’s practice of citing international law and courts in two parts. The first part provides a broad overview of the Court’s practice of citing international law and courts, highlighting the partial and somewhat superficial recourse to international law in much of the Court’s jurisprudence. The second part sets out data demonstrating the Court’s non-citation of the African Court and considers areas where African Court jurisprudence could provide ‘added value’ to the Court’s case-law.

A Overview: the Constitutional Court’s citation of international law and courts

If the first hallmark of the Constitution was the establishment of a powerful domestic Constitutional Court, its second significant hallmark was surely the formal place accorded to international law by the constitutional text. The now-famous s 39 expressly mandates reference by South African courts to international law in interpreting the Bill of Rights, stating:

(1) When interpreting the Bill of Rights, a court, tribunal or forum—
   (a) must promote the values that underlie an open and democratic society based on human dignity, equality and freedom;
   (b) must consider international law; and
   (c) may consider foreign law.62

The main approach, then, was to focus on an ‘inside-out’ approach, where the South African judiciary would reach out to international law norms in interpreting the Bill of Rights, with much less focus on an ‘outside-in’ approach that would entail intervention by international judicial actors: indeed, at the time the Constitution was being drafted, and entered into force, the prospect of an African Court on Human and Peoples’ Rights was yet a mere possibility.

As Dire Tladi has observed, the South African Constitution is ‘reputed to be one of the most international law-friendly constitutions in the world’.63 One can find many statements to this effect in the Court’s jurisprudence. For instance, in the Glenister decision of 2011, Ngcobo J offered:

Our Constitution reveals a clear determination to ensure that the Constitution and South African law are interpreted to comply with international law, in particular international human-rights law.

62 A similar provision can be found in the 1993 Interim Constitution. Section 35(1) stated: ‘In interpreting the provisions of this Chapter a court of law shall promote the values which underlie an open and democratic society based on freedom and equality and shall, where applicable, have regard to public international law applicable to the protection of the rights entrenched in this Chapter, and may have regard to comparable foreign case law.’
These provisions of our Constitution demonstrate that international law has a special place in our law which is carefully defined by the Constitution. Yet Tladi also emphasises that South African courts, including the Constitutional Court, have struggled to set out a sound and systematic methodology for addressing and interpreting international law, with the result that references to international law and adjudication on complex matters of international law can be quite superficial. Similarly, Daniel Abebe has observed, regarding the African Charter on Human and Peoples’ Rights and other international rights treaties:

The South African Constitutional Court has held that they should be directly applicable in South African courts without domestic implementing legislation, but the record is unclear. At times, the Constitutional Court’s recourse to international law has also been viewed as highly instrumental; used in part as a judicial strategy to achieve certain adjudicative ends while attempting to shield the Court from executive opprobrium. In *Glenister*, the Court was called to intervene to stymie legislation affecting the independence of the National Prosecuting Authority (NPA), viewed by the applicants as attenuating the capacity of prosecutorial agencies to address official corruption. In a careful judgment, delivered in March 2011, the Court recognised that transfer of some anticorruption powers to the police and disbandment of a particular anticorruption unit within the NPA were, in principle, permissible, but that the amendment removed important protections of prosecutorial independence by placing power in the hands of political actors who might themselves be subject to prosecution.

As Issacharoff recounts, the *Glenister* Court eschewed the option of basing its judgment on democratic principles within the Constitution. It chose instead to invoke s 39, as well as ss 231(2) and 7(2), to ground its holding that international conventions to which South Africa is a party require member states to maintain anticorruption agencies with a sufficient level of independence, with the result that a failure to meet this requirement could not be considered reasonable. Issacharoff argues that although the reasoning was not entirely convincing, the Court’s recourse to international obligations allowed it to escape a more uncomfortable ruling that challenged the Zuma government head-on:

Placing responsibility for its decision on international law is an interesting judicial expedient. It has the effect of avoiding a direct confrontation with the constitutional underpinnings of democratic authority and instead turning attention to the commands of foreign engagements. The court could sidestep any engagement with the hard questions of the one-party weight of the ANC and instead purport to act as the simple messenger of international law. It was the South African government that entered into the international covenants and the court could act as if its hands were tied.

---

64 *Glenister v President of the Republic of South Africa and Others* [2011] ZACC 6, 2011 (3) SA 347 (CC), 2011 (7) BCLR 651 (CC) para 97.
65 Tladi (note 63 above) at 338.
67 Note 64 above.
69 Ibid at 261. Section 231(2) of the Constitution states: ‘An international agreement binds the Republic only after it has been approved by resolution in both the National Assembly and the National Council of Provinces, unless it is an agreement referred to in subsection (3).’ Section 7(2) of the Constitution states: ‘The state must respect, protect, promote and fulfil the rights in the Bill of Rights.’
70 Ibid at 262.
Glenister sits a little uneasily alongside other decisions such as the Azanian Peoples’ Organisation (AZAPO).\textsuperscript{71} The AZAPO Court emphasised that, while domestic amnesty legislation violated various provisions of international humanitarian law and the right of access to court, the fundamental question was not the legality of amnesty according to international law, but within the terms of the Constitution itself.\textsuperscript{72}

### B Exploring the Constitutional Court’s (non-)citation of African Court jurisprudence

It is against the brief overview above of the Court’s general approach to citing international law and courts that we turn to the Court’s approach to African Court case-law, and the African Charter more broadly.

First, it is worthwhile to note that, unlike the strong and region-wide domestic judicial practice of referring to international human rights law, in both Latin America and Europe, the highest domestic courts across AU Member States refer relatively rarely to international law. Although common law courts appear to show a greater openness than courts in civil-law systems (eg, Chad, Senegal), even within the common-law category there is wide diversity: for instance, the courts of Ghana and Botswana have made use of international law in adjudication, while Zambian courts tend to avoid it.\textsuperscript{73} Of most relevance here, domestic courts tend not to refer to the jurisprudence of the African Court (or other international courts in the AU). As one scholar has recently observed, despite increasing reference to the decisions of the African Commission by national courts, there is

little evidence of the use of the jurisprudence of other regional and sub-regional courts or bodies such as the African Court and the African Children’s Committee. This is perhaps owing to the fact Africa’s supranational courts and tribunals, apart from the African Commission, are relatively young compared to their European counterparts.\textsuperscript{74}

The South African Constitutional Court is no exception in this regard and, if anything, appears more reluctant than many other courts to embrace African Court case-law. As discussed in Part II, in order to assess whether, and how, the Court cites African Court jurisprudence, and the African Charter more generally, this section relied on a relatively simple methodology. Every judgment of the Constitutional Court since 14 June 2013 (the date of the African Court’s first merits judgment) until 25 February 2019 (the time of writing) has been searched on the South African Legal Information Institute (SAFLII) database using four search terms: ‘African Court’, ‘African Charter’, ‘charter’, and ‘Banjul’ (the latter to catch any reference to the African Charter as the ‘Banjul Charter’). This exercise produces three key insights and provokes one broader reflection, as follows.

First, despite the African Court’s growing corpus of case-law, the Constitutional Court in 277 judgments during this period has not yet cited African Court jurisprudence even once.\textsuperscript{75}

No mention of any of the African Court’s 23 merits decisions to date could be found in the

\textsuperscript{71} Azanian Peoples Organization (AZAPO) v President of the Republic of South Africa [1996] ZACC 16, 1996 (8) BCLR 1015, 1996 (4) SA 672 (CC).


\textsuperscript{74} B Dinokopila ‘The Impact of Regional and Sub-regional Courts and Tribunals on Constitutional Adjudication in Africa’ in CM Fombad (ed) Constitutional Adjudication in Africa (2017) at 236.

\textsuperscript{75} The breakdown by year is: 32 (2013); 39 (2014); 40 (2015); 55 (2016); 49 (2017); 53 (2018); 9 (2019).
case-law search. This is, in itself, an unexpected finding. In the earlier years covered by this enquiry – say, from the African Court’s first judgment in June 2013 until the end of 2016 – this is perhaps more understandable, as there was a mere trickle of judgments, and there may have been a time-lag in the Constitutional Court becoming aware of key judgments. However, as the African Court’s volume of jurisprudence has grown this seems less viable as an explanatory factor.

Second, and relatedly, in the almost six-year period covered by this research, only sporadic references to the African Charter can be found, as well as isolated references to other African Union rights instruments, such as the Protocol to the African Charter on the Rights of Women in Africa,76 the African Charter on the Rights and Welfare of the Child,77 and the AU Resolution on Police Reform, Accountability and Civilian Police Oversight in Africa.78 In most cases, reference to the African instrument is rather cursory and no extended interpretation is provided: see, for instance, the fleeting reference to the Charter in the Nkabinde judgment79 concerning the best interests of the child, or regarding the meaning of the phrase ‘freedom, and security of the person’ in AB v Minister of Social Development.80 Rare examples of slightly more extended analysis include reference to Charter art 18 on the family in DE v RH, concerning a spouse’s right of action in delict against a third party for adultery under common law.81

This might be starting to change. In two recent judgments – Gavric v Refugee Status Determination Officer82 (September 2018) and Maledu v Iteleeng Bakgatla Mineral Resources (Pty) Ltd83 (October 2018) – the Court placed significant emphasis on the African Charter, and the African Commission’s landmark decisions in FIDH v Senegal84 and Endorois.85 These cases, respectively, concerned the validity of amnesty for political crimes and the necessity to ensure the free, prior and informed consent of local communities for development and investment projects. It is hard to predict whether the Court’s references to Commission decisions will become more common: it has been observed that since its establishment the Court’s references to, and reliance on, the African Charter and other AU instruments and norms have been rather inconsistent; for instance, placing strong reliance on the Charter in some early decisions but very little reliance in decisions such as the ‘striking’ neglect of the Charter in the Court’s First Certification judgment in 1996.86

---

76 Levenstein and Others v Estate of the Late Sidney Lewis Frankel and Others [2018] ZACC 16, 2018 (8) BCLR 921 (CC) at para 60.
78 McBride v Minister of Police and Another [2016] ZACC 30; 2016 (2) SACR 585 (CC), 2016 (11) BCLR 1398 (CC) at para 34.
80 AB and Another v Minister of Social Development [2016] ZACC 43, 2017 (3) BCLR 267 (CC), 2017 (3) SA 570 (CC) para 309.
82 [2018] ZACC 38, 2019 (1) SA 21 (CC), 2019 (1) BCLR 1 (CC).
85 Centre for Minority Rights Development (Kenya) and Minority Rights Group International obo Endorois Welfare Council v Kenya 2009 AHRLR 75 (ACHPR 2009) (‘Endorois’).
Third, in contrast to the complete absence of references to African Court case-law, the Court appears far more open to citing other international courts during this period, especially the European Court of Human Rights (ECtHR) and the Court of Justice of the European Union (CJEU). In the five-year period covered, one finds a raft of references to the ECtHR, and various references to the CJEU. This is alongside citation of customary international law, United Nations standards, Council of Europe standards – and of leading charters and national courts worldwide, especially the Canadian Supreme Court and the Canadian Charter of Rights and Freedoms. This, it must be said, has been a feature of the Constitutional Court’s jurisprudence since its founding: in the first two years of its operation, for instance, we find very regular and substantial references to the Canadian Charter and Canadian Supreme Court case-law compared to a mere handful of fleeting references to the African Charter.

The broader reflection these three key insights prompt is what, if anything, is missing from the Court’s case-law due to its failure to cite African Court jurisprudence, or the African Charter more broadly. For some cases, it appears evident that African Court jurisprudence is largely irrelevant: in cases concerning company law, trusts, tenancy law, employment law, and succession law, for instance.

However, in other cases existing African Court case-law appears highly relevant. Two examples will suffice here. First, in Democratic Alliance v African National Congress, decided on 19 July 2015, the Constitutional Court addressed the right to freedom of expression and

---

87 De Vos NO and Others v Minister of Justice and Constitutional Development and Others [2015] ZACC 21, 2015 (2) SACR 217 (CC), 2015 (9) BCLR 1026 (CC) at paras 22 and 37; Shoprite Checkers (Pty) Limited v Member of the Executive Council for Economic Development, Environmental Affairs and Tourism, Eastern Cape and Others [2015] ZACC 23, 2015 (6) SA 125 (CC), 2015 (9) BCLR 1052 (CC) at paras 18, 63, 111 and 153; Legal Aid South Africa v Magidiwana and Others [2015] ZACC 28, 2015 (6) SA 494 (CC), 2015 (11) BCLR 1346 (CC) at para 111; Kham and Others v Electoral Commission and Another [2015] ZACC 37, 2016 (2) BCLR 157 (CC), 2016 (2) SA 338 (CC) at para 84; Nkabinde and Another v Judicial Service Commission and Others [2016] ZACC 25, 2016 (11) BCLR 1429 (CC), 2017 (3) SA 119 (CC) at para 8; Hotz and Others v University of Cape Town [2017] ZACC 10, 2017 (7) BCLR 815 (CC), 2018 (1) SA 369 (CC) at para 30; Saidi and Others v Minister of Home Affairs and Others [2018] ZACC 9, 2018 (7) BCLR 856 (CC) at para 32.


89 Saidi v Minister of Home Affairs (note 87 above).

90 De Vos v Minister of Justice (note 87 above) at para 29 (UN Convention on the Rights of Persons with Disabilities) and and My Vote Counts NPC v Speaker of the National Assembly and Others [2015] ZACC 31 at para 15 (UN Convention Against Corruption).

91 McBride v Minister of Police (note 78 above) at para 34 (Council of Europe’s Commissioner for Human Rights’ Opinion on the Independent and Effective Determination of Complaints Against the Police).

92 AB v Minister of Social Development (note 80 above) paras 136 and 305; and Dladla and Another v City of Johannesburg and Others [2017] ZACC 42, 2018 (2) BCLR 119 (CC), 2018 (2) SA 327 (CC) footnotes to paras 98 and 99.


the right to vote in free and fair elections, in a case concerning the Democratic Alliance (DA)’s issuance of a short messaging service (SMS) to 1.5 million voters concerning then President Jacob Zuma and the Nkandla Report (on the President’s corrupt use of public monies) ahead of the 2014 elections. However, despite its clear salience, the Court makes no reference in its decision to the African Court’s existing judgments, including its landmark judgment in *Mtikila v Tanzania* over eighteen months earlier, concerning the right to political participation, nor its free speech judgments in the *Zongo* and *Konaté* cases against Burkina Faso. Nor does the Court make any reference to the relevant rights in the African Charter (art 9(2) right to freedom of expression and art 13 right to political participation).

Specific aspects of the African Court’s case-law resonate with the Constitutional Court’s judgment. For instance, concerning the ANC’s allegation that the Democratic Alliance party’s SMS constituted a breach of the Electoral Act and/or the Electoral Code of Conduct issued under that Act, the joint judgment of Cameron, Froneman and Khampepe JJ (with Mosenek DCJ and Nkabinde J concurring) emphasised that penal provisions should be interpreted restrictively, on the basis that ‘freedom of expression is the cornerstone of democracy’, and as a longstanding principle of the rule of law, the common law and of the Court’s own case-law. This chimes with the African Court’s holding in *Mtikila* that the ‘claw-back’ clauses in the African Charter, which textually provide a wide basis for rights restriction, should not be interpreted against the Charter and that regulation of rights and freedoms ‘may not be allowed to nullify the very rights and liberties they are to regulate’, and its judgment in *Konaté v Burkina Faso*, holding that restrictions on freedom of expression must be proportionate, and noting the specific rule laid down in the African Commission’s Declaration of Principles on Freedom of Expression that ‘sanctions should never be so severe as to interfere with the exercise of the right to freedom of expression’.

Contextual factors further amplifying the relevance of the African Court’s judgment in *Mtikila* include the fact that this case, at its core, centred on the disentrenchment of one-party rule and entrenchment of a multiparty system in Tanzania, with the African Court placing emphasis on the need for the electorate to be accorded sufficient choice, questioning whether sufficient choice existed ‘if in order to even choose a representative of one’s choice one is compelled to choose only from persons sponsored by political parties, however unsuitable such persons might be’. This is an issue of clear resonance for South Africa’s dominant-party system. In *Democratic Alliance* the Court not only began by noting that the applicant was the official opposition party in Parliament, but the joint judgment of Cameron J et al similarly emphasised the structural basis for according generous free speech rights within the electoral context, stating: ‘The right individuals enjoy to make political choices is made more meaningful by challenging, vigorous and fractious debate’.

These are just quick illustrations of relevant aspects of the African Court’s case-law, which would also have had potential relevance in other cases such as *Kham v Electoral Commission*, decided in November 2015, which concerned the Electoral Commission’s duty to register

---

95 Ibid at paras 121–122, 129.
96 Ibid at para 109.
97 Ibid at paras 125–150, 151.
98 Ibid at para 109.
99 Ibid at paras 6, 135.
100 *Kham and Others v Electoral Commission and Another* [2015] ZACC 37, 2016 (2) SA 338 (CC).
voters in the correct voting district, and *Maledu*,\(^\text{101}\) discussed above, where the Court cited the African Commission decision in the *Enderois* case, but not the African Court judgment in *Ogiek*, in a case concerning free, prior and informed consent for development projects. Even in some cases that appear entirely rooted within the particularities of the South African historical, social, political and constitutional context – such as cases concerning the legacy of apartheid\(^\text{102}\) – reference to African Court jurisprudence and the African Charter on Human and Peoples’ Rights could enrich the analysis and place it in a wider context. After all, the preamble to the African Charter expressly refers to the Charter of the Organization of African Unity (OAU) statement that ‘freedom, equality, justice and dignity are essential objectives for the achievement of the legitimate aspirations of the African peoples’ and refers to the need to eliminate apartheid (as well as colonialism, neocolonialism and other forms of domination). This may seem itself superficial, but it is worthwhile to emphasise that there is no comparable statement in any other international human rights treaty.

This is not necessarily to make the argument for a generalised citation of African Court jurisprudence, and of the African Charter itself, or for citing the African Court in cases that can be decided within the four corners of the domestic Constitution. Rather, it is to at least raise the point that in many instances these sources might enrich the South African Constitutional Court’s jurisprudence and anchor it more firmly within the developing regional system of human rights protection, adding to an already well-established practice of analysing foreign and international norms and jurisprudence. In some cases, as seen above, clearly relevant African Court decisions have not been cited. What accounts for this? Is the Constitutional Court snubbing its regional counterpart, or is there a wider array of explanatory factors at play?

V \hspace{1em} IS THE CONSTITUTIONAL COURT SNUBBING THE AFRICAN COURT? ‘IT’S COMPLICATED’

At one level, the Constitutional Court’s non-citation of African Court jurisprudence could be approached as simply one of institutional preference – or individual judicial preference which happens to be shared by all eleven judges across the Court. One could also frame it as an issue of supremacy: having accreted an appreciable level of hard-won constitutional supremacy since the mid-1990s, the Constitutional Court may be unwilling to cede a share to an international court, or may fear losing a significant level of adjudicative autonomy if it tethers itself too closely to its regional counterpart.\(^\text{103}\) We might also characterise it as just one dimension of a general propensity to cite case-law from outside Africa: for instance, regarding citation of foreign courts (rather than international law and courts) Joseph Isanga suggests that the Constitutional Court’s tendency to cite US and European courts reflects a more general tendency – seen also in other ‘successful’ courts in Botswana, Ghana, Malawi, and Namibia – to rely unduly on ‘non-African jurisprudence’ to validate judgments.\(^\text{104}\)

\(^{101}\) *Maledu* (note 83 above).

\(^{102}\) *City of Tshwane Metropolitan Municipality v Afriforum and Another* [2016] ZACC 19, 2016 (9) BCLR 1133 (CC), 2016 (6) SA 279 (CC); *South African Revenue Service v Commission for Conciliation, Mediation and Arbitration and Others* [2016] ZACC 38, [2017] 1 BLLR 8 (CC), (2017) 38 ILJ 97 (CC), 2017 (1) SA 549 (CC), 2017 (2) BCLR 241 (CC); *Dladla and Another v City of Johannesburg and Others* [2017] ZACC 42, 2018 (2) BCLR 119 (CC), 2018 (2) SA 327 (CC); and *AfriForum and Another v University of the Free State* [2017] ZACC 48, 2018 (2) SA 185 (CC), 2018 (4) BCLR 387 (CC).

\(^{103}\) Roux (note 16 above) (On the Constitutional Court’s hard-won authority).

\(^{104}\) Isanga (note 28 above) at 752, fn 24.
Constitutional Court judges, such as Justice Sachs, have expressed support for comparative African jurisprudence, Isanga states: ‘the South African Constitutional Court has referenced more non-African jurisprudence than African jurisprudence in its judicial review.’ It is important here to acknowledge that this appears often related to the limited availability of, and familiarity with, jurisprudence from other African states, including the limited availability of secondary sources to contextualise which judgments are accessible and fully understand their place in the case-law of courts in other African states.

A range of additional explanatory factors can also be considered for the Court’s failure to cite the African Court, which suggest that responsibility cannot be laid entirely at the Court’s door. This section canvasses seven such factors, ranging from macro-political factors, to strategic institutional factors, to broad structural factors.

First, the broad macro-political environment does not incentivise the Court to look to the regional level. It is important to recall that the Organisation of African Unity (OAU) was not replaced by the African Union until 2002 and judicial or quasi-judicial mechanisms at the regional level had made little impact at the national level as the Constitutional Court worked out its role in the new democratic dispensation. The African Commission on Human and Peoples’ Rights, for instance, created as a stand-alone institution in 1987, and faced with a more challenging context of post-colonial states running the gamut from authoritarianism to tentatively consolidating democracy, found little room to manoeuvre in its first decades. It adopted a more deferential posture to states than its counterparts in other regions, through a focus on ‘positive dialogue’, inconsistent use of provisional measures, and reluctance to follow up its decisions, and at the time the South African Constitutional Court was established the Commission had yet to issue its most assertive decisions.

More fundamentally, it may be offered that the ANC, in the grand political settlement underlying South Africa’s transition from minority rule under apartheid to majority rule under the new democratic dispensation, had submitted to a very particular form of domestic judicial power – embodied in the Constitutional Court. It did not submit to judicial power in any form. In addition, the ANC was eager to place the state within the mainstream of international law, to end South Africa’s status as a pariah state in the international community under apartheid, as evidenced in its ratification of a raft of international human rights treaties throughout the 1990s (eg, the International Covenant on Civil and Political Rights, the International Convention on the Elimination of Racial Discrimination). However, post-apartheid South Africa was, as Peter Vale puts it, a ‘reluctant regionalist’ in economic affairs, which may also explain its approach to regional human rights protection, an issue discussed in more depth below:

105 Ibid.
106 I am grateful to Advocate Michael Bishop for this insight. He further notes that while there is interesting jurisprudence coming from Kenya, Tanzania, Uganda, Nigeria and other states, secondary materials on this case-law would more readily facilitate its citation. Advocate Bishop observes that a recent collection of translated and annotated excerpts of the Colombian Constitutional Court’s judgments has greatly increased his ability to draw down upon Colombian case-law. M Cepeda Espinosa & D Landau *Colombian Constitutional Law: Leading Cases* (2017).
107 The only states considered by Freedom House to be electoral democracies in the late 1980s were Botswana, the Gambia, and Mauritius.
109 A useful list of treaty ratifications is provided by the University of Minnesota’s Human Rights Library, available at http://bit.ly/2t9tEZw.
For all the pageantry, pomp and pronouncements of South Africa’s new place in the order of regional things, the country was a reluctant regionalist. Not only were the bureaucrats responsible for making the first links into the region’s multilateralism drawn from the country’s apartheid past, but economic discourse within South Africa had turned its attention away from the region. As a 1994 report issued by the African Development Bank noted: ‘What is clear is that for South Africa national interests are paramount, while regional issues are secondary and likely to remain so.’ This emphasis on South Africa’s own interests, rather than on developing a common regional purpose, ended any hope that the region could become more than the sum of its separate sovereign pieces.110

The almost exclusive focus on domestic counter-majoritarian institutions in South Africa’s democratic transition lies in significant contrast to democratic transitions in Central and Eastern Europe and South America throughout the 1980s and 1990s. In the latter transitions, the sweeping region-wide shift from authoritarianism to democratic rule provided a sound basis for action by a regional human rights court. In Europe it required the re-making of the European Court of Human Rights as an aid to prevent the re-emergence of totalitarian regimes. In South America it provided the Inter-American Court of Human Rights with the space to carve out a rule in assisting pushback against reconsolidation of military governments or extreme right-wing regimes with strong ties to the military. The African scenario was different. Despite a much greater focus on democratic rule and constitutionalism in the 1990s, there was no sweeping or universal region-wide democratic transformation.111 South Africa, albeit a totemic and era-defining transition to democracy, was part of a much patchier and more atomised set of African democratic transitions during the 1990s, with the result that, although it was a highly internationalised process, it was not a regionalised process.

Second, the fact that the South African government has not made the optional declaration to permit individual and NGO petitions to the African Court may be viewed as a barrier to freely citing the latter’s case-law. However, such an argument (if entertained within the Constitutional Court) does not hold up when one considers the Constitutional Court’s liberal citation of other courts, even those interpreting normative instruments to which South Africa is not a party (such as the European Convention on Human Rights and EU law).

Third, there may very well be a sense that the successes of the Constitutional Court render an international human rights court obsolete in the South African context. As Andreas O’Shea has observed, various arguments had been made to this effect before the African Court was established:112

This argument rests on the premise that there are adequate mechanisms for the protection of human rights on a national level. It may be said that at national level a constitution with a bill of rights exists. That bill of rights reflects all the important provisions of human rights treaties and may be enforced through a constitutional court that will give primacy to the constitution and the bill of rights. What need is there then for yet another body to perform this identical judicial function? The South African Constitutional Court may serve as an example. The Constitutional Court applies the Constitution that incorporates most of the content of the African Charter and arguably goes further. Other decisions, rulings and legislation may be declared unconstitutional if

they infringe the Bill of Rights. The Court itself operates in a very similar fashion to the proposed African Court. It consists, like the African court, of 11 judges, its decisions are final and binding and its judges are in practice selected from personalities that have struggled for the protection of human rights and fundamental freedoms.

Fourth, a number of rational strategic considerations – beyond mere preference, discussed above – may inform the Constitutional Court’s approach to citation of international law. If a central aim of such citation is to bolster the authority of its jurisprudence, it is understandable that the Court would cite more venerable courts such as the US Supreme Court, the Canadian Supreme Court, and the ECtHR, rather than younger courts. The objective of shielding the Constitutional Court may have become more acute in recent years, as it has weathered periodic attacks from the Zuma administration, including the government’s announcement of a review of the Court’s powers in 2012.113

Fifth, seniority or vintage may play a broader part here. From a comparative perspective, it is notable that the national courts that have evinced most resistance (even if mainly principled) to the regional human rights court in other world regions were all established before the regional court, ie the greatest challenges to the ECtHR, which began functioning in 1959, has come from the constitutional courts of Germany (1951) and Italy (1956), and the UK Supreme Court (successor of the centuries-old Judicial Committee of the House of Lords). The Brazilian Supreme Court (first established in the republican Constitution of 1891) has taken a distinctly frosty attitude toward the IACtHR, which began operating in 1979.114 The fact that the South African Constitutional Court was established twenty years before its regional peer may be significant. More broadly, the ‘judicial maturity’ of the African regional courts system as a whole has been cited by Daniel Abebe as a key factor hindering its impact at the domestic level.115

Sixth – and building on the lack of incentives to cite the African Court – it appears that there is no countervailing force pushing the Constitutional Court to cite African Court jurisprudence. It is clear from the case-law review in part IV, above, that in many cases it is the applicants’ submissions that direct the Court toward specific sources of international law.116 In this respect, it is notable that detailed knowledge of the African Court in South Africa – and indeed, across the African Union – remains minimal. As the Court’s former president observed, even in the Court’s permanent seat, the city of Arusha in northern Tanzania, ‘there are people who are wondering if there is such a court in the city.’117 A 2015 interview with Lenser Anyango of the Network of African National Human Rights Institutions (NANHRI) – which brings together 44 national human rights institutions from across the region (including the South African Human Rights Commission (SAHRC) – revealed a strong sense among

---


115 Abebe (note 66 above) at 572–573.

116 Rural Maintenance (Pty) Limited and Another v Maluti-A-Phofung Local Municipality [2016] ZACC 37, 2017 (1) BCLR 64 (CC) at paras 2, 22.

human rights activists across the continent that the Court is an ‘alien institution’ and that most
domestic rights bodies are ‘detached, disinterested and disconnected from the African Court process.” 118
It is likely, then, that applicants are not citing African Court jurisprudence or the African
Charter when petitioning the South African Constitutional Court. That said, limited evidence
exists that even when applicants have included reference to African Court case-law in their
submissions to other superior courts in South Africa, these references do not make their way
into the Court’s judgment. 119 Of course, it is possible that the Constitutional Court does
analyse African Court jurisprudence without explicitly citing it in its judgments: this practice
is found in a variety of courts. 120
Seventh, and finally, other practical considerations may also be at play: judges on the
Constitutional Court may have limited familiarity with the African Court’s case-law; links
between the two courts may be underdeveloped (despite the biennial African Judicial Dialogue
organised by the African Court, other visits, and plans for an African Judicial Network 121);
and there is currently no South African judge on the African Court (the only South African
judge, Bernard Makgabo Ngoepe J, served two terms, a two-year term and six-year term, from
2006–2014). This may all compound the common lack of in-depth knowledge across domestic
courts, concerning the African Court and how it operates. As Abebe notes:

Whether the African court system could catalyze improved human rights enforcement by
domestic courts is an open question, but the lack of knowledge about the [African Commission
on Human and Peoples’ Rights] and its operations by domestic courts, for example, is not
promising evidence. 122

VI CONCLUSION: WHY DOES THIS RELATIONSHIP MATTER?
This article has picked over the odd relationship between the South African Constitutional
Court and the African Court on Human and Peoples’ Rights, attempting to divine the reasons
why a national court with such a strong affinity – in principle – to its regional counterpart
would ignore the latter’s jurisprudence. While it is not possible to say with any certainty why
this is the case, by canvassing a variety of possible explanations, and considering the question
from a comparative perspective, the aim of this article was to throw some light on a dimly-lit
area of research on the much-studied Constitutional Court.

This relationship matters. The South African Constitutional Court’s lack of engagement
with African Court jurisprudence means that it may possibly be foregoing an opportunity to
genuinely enrich its own case-law, and to anchor it more firmly in the developing pan-regional

118 ‘An Alien Institution: A Q&A with the Network of African National Human Rights Institutions’ The ACtHPR
119 I am grateful to Advocate Michael Bishop for this insight. He had cited the African Court’s Ogiek judgment
in heads of argument submitted to the Supreme Court of Appeal. Ogiek does not appear in the Court’s final
120 For instance, at an event at the Honorable Society of King’s Inns (Ireland) on 12 February 2019 marking the
15th anniversary of the Human Rights Act 2003, the Chief Justice of Ireland remarked that while ECtHR
jurisprudence is not always addressed in written judgments at the domestic level, it may influence outcomes.
This unrecorded but palpable influence affects the decisions of the Taiwanese Constitutional Court. D Law &
ly/2Cmep65.
122 Abebe (note 66 above) at 570.
human rights system. Moreover, as the African Court’s jurisprudence grows, failure by the Constitutional Court to cite it will seem increasingly incongruous; what now seems curious will start to seem like a deafening silence. There is no doubt that the African Court’s case-law is set to rapidly expand in the near future: its website lists over 100 cases pending before the Court (admittedly, 80 of these concern the Court’s host state, Tanzania).

More importantly, it appears that the South African Constitutional Court is in a uniquely influential position – as against its national peers – to provide support to the development of the African Court as a key site for the elaboration of a transregional epistemic community centred on the African Charter of Human and Peoples’ Rights; the most widely ratified rights treaty in the African Union, and the primary instrument capable of providing a focal point for national and international courts across the AU. Were the Court to invest energy in developing familiarity with, and a practice of citing, African Court jurisprudence, it would likely encourage other domestic courts to do the same. The African Court, for its part, could also engage more readily with Constitutional Court case-law.

The argument, it should be emphasised again, is not for citation and support from the South African Court as an act of judicial courtesy or even judicial charity, nor for artificial recourse to African Court case-law when it is not directly relevant. Rather, as perhaps the two most important courts on the entire continent as regards rights protection, it seems desirable that their relationship should be developed and deepened. Evidently, this does not mean they will always agree, but communication is preferable to the development of parallel lines of jurisprudence in isolation from one another. Only time will tell if the kindred strangers can become kindred spirits.
Human Dignity in the Common Law of Contract: Making Sense of the Barkhuizen, Bredenkamp and Botha Trilogy

HANRI DU PLESSIS

ABSTRACT: This article considers the role of human dignity in the South African common law of contract. Firstly, I critically investigate how human dignity is understood in contract law. This investigation consists of a discussion of Beyleveld and Brownsword’s dualistic approach to human dignity based on their individualistic interpretation of Kantian dignity. As an alternative to this individualistic reading of Kant, I then discuss Cornell and Wood’s communitarian interpretation of Kantian dignity. Thereafter, I rely on Cornell’s criticisms of Kantian dignity to highlight and compare the notion of ubuntu with the Kantian interpretations of dignity. I rely on the work of Cornell (together with Muvangua) to investigate human dignity through ubuntu in order to show how Kantian dignity and ubuntu can be harmonised. In the second part of the article, I discuss the Court’s approach to human dignity as developed in the constitutional jurisprudence outside the field of contract law over the last two decades. The Court’s approach reflects a similar harmonisation between Kantian dignity and ubuntu. I then use the Court’s evolving understanding of the constitutional value of human dignity to critically engage with the apparent tension between the Court’s approach to contractual fairness in Barkhuizen v Napier and Botha v Rich NO, with that of the Supreme Court of Appeal in Bredenkamp v Standard Bank of South Africa Ltd. I conclude by arguing that the decision in Botha can be interpreted as creating an equitable discretion to declare unfair contract terms or the unfair enforcement of contract terms invalid and that the Court is in the process of developing an equitable discretion in the common law of contract.

KEYWORDS: contractual fairness, human dignity as empowerment, human dignity as constraint, Kantian dignity, ubuntu

AUTHOR: Senior Lecturer, Department of Private Law, University of South Africa. Email: dplesh@unisa.ac.za

ACKNOWLEDGEMENTS: The article forms part of research undertaken during a period of research and development leave granted by Unisa in 2019. Parts of this article are also based on research forming part of my unpublished thesis entitled The Harmonisation of Good Faith and Ubuntu in the South African Common Law of Contract (LLD, Unisa, 2017). The research for my thesis was funded by a grant from Unisa’s
Academic Qualification Improvement Programme from 2014–2017. Any opinions or conclusions expressed in this article are my own and Unisa does not accept any liability for them. Similarly, all errors remain my own. I would like to thank the following persons for their valuable comments on previous drafts of this article: Prof Stu Woolman, Advocate Martin Brassey SC and the two anonymous reviewers.
I INTRODUCTION

The role of fairness in the South African common law of contract has been a contentious issue for centuries. After the introduction of the Constitution of the Republic of South Africa, 1996 (Constitution), this debate came to the fore with the apparent tension between the Court’s decision in Barkhuizen and that of the Supreme Court of Appeal in Bredenkamp. The Barkhuizen Court expressly rejected the contention that a contractual term that is unfair cannot lead to the conclusion that the term is contrary to constitutional values and principles. In contrast, the Bredenkamp court held that fairness becomes relevant only when a specific constitutional value or right is implicated and it must be determined whether the clause or its enforcement is contrary to public policy. Accordingly, it rejected the idea that fairness is an overarching requirement in the common law of contract.

Legal practitioners and academics awaited the Court’s response to Bredenkamp, especially after the Everfresh Court’s obiter dictum that ubuntu may require the development of good faith to ensure more contractual justice. However, the Court’s subsequent decision in Botha was met with disappointment. The Court held that the enforcement of a cancellation clause which would also entail the forfeiture of more than half of the purchase price already paid by the purchaser ‘would be a disproportionate penalty for the breach’ of the contract. Some legal scholars argued that the Court’s reasoning led to the impression that it impliedly held that the enforcement of the cancellation clause would be unfair and therefore unconstitutional. Sharrock argued that if this is the case, the Court’s decision is in conflict with that in Bredenkamp, which held that fairness is not an overarching requirement for the enforcement

1 I use the term ‘fairness’ to denote the amorphous concept implied in legal principles ranging from good faith to equity intended to realise justice in contract. This article is premised on the idea that these terms are open norms ie rules or standards that have no fixed or restricted meaning, can apply to various situations and enable value judgments. L Hawthorne ‘Public Policy: The Origin of a General Clause in the South African Law of Contract’ (2013) 19(2) Fundamina 300, 300‒301; D Bhana & M Pieterse ‘Towards a Reconciliation of Contract Law and Constitutional Values: Brisley and Afrox Revisited’ (2005) 122 South African Law Journal 865, 868.
5 Barkhuizen (note 3 above) at para 72.
6 Bredenkamp (note 4 above) at para 43–44.
7 Ibid at paras 50–53. See also Mohamed’s Leisure Holdings (Pty) Ltd v Southern Sun Hotel Interests (Pty) Ltd [2017] ZASCA 176, 2018 (2) SA 314 (SCA) at para 30.
8 Everfresh Market Virginia (Pty) Ltd v Shoprite Checkers (Pty) Ltd [2011] ZACC 30, 2012 (1) SA 256 (CC) (‘Everfresh’) at paras 71–72 (Mosemeke J) and at para 23 (Yacoob J).
9 Botha & Another v Rich NO & Others [2014] ZACC 11, 2014 (4) SA 124 (CC)(‘Botha’).
10 Ibid at para 51 read with paras 45–46.
of a contract.\textsuperscript{12} Bhana and Meerkotter have maintained that as the Court did not explicitly implicate an enumerated right it appears unnecessary to implicate a specific right or value in the Bill of Rights as envisaged by the \textit{Bredenkamp} court.\textsuperscript{13} As the \textit{Botha} Court made no express pronouncements on these issues, the legal position remains uncertain.

In this article, I use the Court’s understanding of the constitutional value of human dignity as developed over the last two decades to critically engage with these three decisions. I argue that the decision in \textit{Botha} can be interpreted as creating an equitable discretion to declare contractual terms to be unfair or declare the unfair enforcement of contractual terms to be invalid by relying of the constitutional value of human dignity, which is based on a harmonisation between Kantian dignity and ubuntu.\textsuperscript{14} For this purpose, the article is divided into the following parts: First, I explain how human dignity is understood in contract law. I start with Beyleveld and Brownword’s dualistic approach to human dignity as based on their individualistic interpretation of Kantian dignity. I then briefly discuss Cornell and Wood’s communitarian interpretation of Kantian dignity as an alternative reading of Kant. Thereafter, I rely on Cornell’s criticisms of Kantian dignity to bring a comparison with ubuntu to the fore. I rely on the work of Cornell (together with Muvangua) to investigate human dignity through ubuntu and show how Kantian dignity and ubuntu can be harmonised. In the second part of the article, I discuss the Court’s approach to human dignity as reflected in constitutional jurisprudence outside the field of contract law which is based on harmonisation between Kantian dignity and ubuntu. Finally, I use the Court’s evolving understanding of human dignity and other contract law decisions to critically engage with \textit{Barkhuizen, Bredenkamp} and \textit{Botha}, and argue that the Court is in the process of developing an equitable discretion in the common law of contract.

\section*{II APPROACH TO HUMAN DIGNITY IN CONTRACT LAW}

As human dignity is an open-ended\textsuperscript{15} concept, determining its meaning is not an easy task.\textsuperscript{16} Its meaning depends on the specific political and/or moral philosophy used to inform its content.\textsuperscript{17} Feldman explains that it is generally viewed as either serving individual or collective interests.\textsuperscript{18} The same problem is faced when defining the constitutional value of human dignity

\begin{itemize}
  \item \textsuperscript{12} Sharrock (note 11 above) at 179–183.
  \item \textsuperscript{13} Bhana & Meerkotter (note 11 above) at 505.
  \item \textsuperscript{14} I accept that the constitutional value of the rule of law and the public policy consideration of legal certainty are relevant when dealing with contractual fairness but this topic falls outside the scope of this article. Du Plessis (note 2 above) at 328–333 read with 77–82 (where I argue that an equitable discretion should not necessarily lead to large scale legal and commercial uncertainty if used cautiously, in an incremental manner and within the existing legal principles as far as possible.)
  \item \textsuperscript{15} Compare the description of open-ended terms (note 1 above).
  \item \textsuperscript{16} \textit{Teddy Bear Clinic for Abused Children & Another v Minister of Justice and Constitutional Development & Another} [2013] ZACC 35, 2014 (2) SA 168 (CC)(‘\textit{Teddy Bear Clinic}’) para 52.
  \item \textsuperscript{17} D Feldman ‘Human Dignity as a Legal Value – Part 1’ (1999) \textit{Public Law} 682, 686.
  \item \textsuperscript{18} Ibid at 685. See also C Albertyn ‘Values in the South African Constitution’ in D Davis, A Richter & C Saunders (eds) \textit{An Inquiry into the Existence of Global Values Through the Lens of Comparative Constitutional Law} (2015) 321, 326.
\end{itemize}
in the South African common law of contract. Therefore, before investigating the meaning of the constitutional value of human dignity in the South African common law of contract, it is necessary to investigate how the dualistic approach (including its philosophical origins) is understood in contract law.

A Dualistic approach to human dignity in contract law

South African contract law scholars generally refer to Beyleveld and Brownsword’s work when discussing the role of human dignity in contract law. This necessitates a critical investigation of their work. These authors distinguish between two concepts of human dignity, namely empowerment and constraint, and they trace both concepts back to Kantian dignity. Human dignity as empowerment refers to the case where human dignity promotes individual interests in contract law, while human dignity as constraint promotes communitarian interests.

1 Human dignity as empowerment

This conception of human dignity entails the empowerment of a person to live an autonomous life by making her own decisions and taking responsibility for those decisions. Brownsword provides the following definition:

[I]t is because humans have a distinctive value (their intrinsic dignity) that they have rights qua humans. Commonly, it is the capacity for autonomous action that is equated with human dignity and this, in turn, generates a regime of human rights organised around the protection of individual autonomy.

Brownsword and Beyleveld argue that the intrinsic worth of all human beings, and therefore, their dignity, finds support in Kantian ethics. Kant argued that something has a dignity because it has intrinsic value and is of inestimable worth. Kant also maintained that ‘the dignity of man consists precisely in his capacity to make universal law, although only on condition of being himself also subject to the law he makes’. Relying on this passage,

---


20 Du Plessis (note 2 above) at 215–235.

21 Hawthorne (note 19 above) at para 4 1; Barnard (note 19 above) at 231–232; Lubbe (note 19 above) at 420–421; Bhana & Pieterse (note 1 above) at 881.


23 Ibid.

24 Feldman (note 17 above) at 685.


26 Ibid at 191.


28 Ibid at 101.
Brownword and Beyleveld propose that Kantian dignity can explain why human beings have dignity; and supports human dignity as empowerment that is based on human beings’ inherent worth, the latter being founded on their capacity for autonomous action.

Brownword and Beyleveld state that human dignity as empowerment requires that human beings should be treated as autonomous beings who can make their own decisions and should never be treated as mere things or instruments. They trace this idea back to Kant’s second categorical imperative which provides that one should ‘treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end’. They rely on this passage to support their argument ‘that it is wrong to treat persons as mere things rather than as autonomous ends’. Their interpretation of Kant therefore supports their empowerment conception of human dignity which requires the protection of individuals’ ‘choices against the unwilled interferences of others’.

According to Brownword, human dignity as empowerment comprises two aspects: firstly, protection against direct attacks, for example, slavery. Secondly, the creation of the necessary conditions for the realisation of a person’s human dignity, which translates as a ‘right to support and assistance to secure circumstances that are essential if one is to flourish as a human’. An example of the latter is the various socio-economic rights that must be protected and promoted by the state.

This conception of human dignity is rights-driven because it grounds ‘a set of rights claims against others’ and reinforces ‘claims to self-determination’. It places great importance on individual autonomy and maximises individual freedom which may only be limited to protect the rights of others ‘or for some less direct rights-related reason’. Hence it is based on the political philosophies of individualism and economic liberalism which limits state interference in private transactions.

As applied in a contract law setting it means that contracting parties are free to conclude contracts as an expression of their autonomy and that contracts freely entered into must be respected. Hence this notion of human dignity finds expression in the classical law of contract which can also be linked to the political philosophies of individualism and economic liberalism.

---

30 Ibid at 666.
31 Ibid.
32 Kant (note 27 above) at 91 as quoted by Beyleveld & Brownword (note 29 above) at 666.
33 Ibid.
34 Brownword (note 25 above) at 183.
37 Beyleveld & Brownword (note 22 above) at 27–28.
39 Ibid at 189.
40 Brownword (note 25 above) at 193–194.
41 Ibid at 184–185.
2 Human dignity as constraint

Human dignity as constraint refers to the limitation of individual autonomy where the protection and promotion of another’s human dignity requires it.\(^{42}\) It expresses a duty-driven approach to human rights\(^{43}\) and denotes the correlating duty to respect the human dignity of another as well as the duty to respect another’s human dignity as expressed through her specific human rights.\(^{44}\) This conception of human dignity has been linked to Kant’s duty not to ‘act contrary to the equally necessary self-esteem of others, as human beings’ and the ‘obligation to acknowledge in a practical way, the dignity of humanity in every other human being’.\(^{45}\)

It may also recognise a further indirect duty to respect the community’s ‘vision’ of human dignity.\(^{46}\) This vision may arise ‘where there is a background ethic of care and concern for others (and, concomitantly, a sense of solidarity with and responsibility for others)’ which results not only in a concern for one’s own flourishing but also that of others.\(^{47}\) In this context it may denote a duty on individuals to assist in the promotion of the necessary conditions for the realisation of others’ human dignity.\(^{48}\) Therefore, human dignity as constraint provides justification for the promotion of socio-economic rights and substantive equality in the private sphere.\(^{49}\)

This conception of human dignity allows for a greater limitation of individual freedom\(^{50}\) and envisages a paternalistic society\(^{51}\) in which the exercise of individual autonomy may be tempered if it is incompatible with the human dignity of others, or in order to create the necessary conditions for the realisation thereof. In other words, private individuals (and not only the state) are obliged to assist in the promotion of substantive equality and an individual’s autonomy may be limited to promote social justice.

In contract law this means that even freely concluded contracts must be tested for conformity to human dignity.\(^{52}\) Hence human dignity as constraint aligns with the socialist values expressed in modern contract law, and permits state intervention in contracts to protect and promote human dignity.\(^{53}\)

3 Tension resulting from the dualistic approach to human dignity

In Beyleveld and Brownsword’s paradigm, the two notions of human dignity are in tension with each other as they pull in opposite directions.\(^{54}\) Human dignity as empowerment supports contractual autonomy while human dignity as constraint results in its limitation.

\(^{42}\) Brownsword (note 35 above) at 1.
\(^{43}\) Beyleveld & Brownsword (note 22 above) at 36.
\(^{44}\) Ibid at 37.
\(^{46}\) Beyleveld & Brownsword (note 22 above) at 37.
\(^{47}\) Ibid at 41–42.
\(^{48}\) A Clapham Human Rights in the Private Sphere (1993) 148. See also Brownsword (note 35 above) at 4.
\(^{49}\) Botha (note 36 above) at 174.
\(^{50}\) Brownsword (note 38 above) at 297–298.
\(^{51}\) Beyleveld & Brownsword (note 22 above) at 11 and 31, esp 4.
\(^{52}\) Brownsword (note 25 above) at 194.
\(^{53}\) Bhana & Pieterse (note 1 above) at 881 and Lubbe (note 19 above) at 422.
\(^{54}\) Brownsword (note 25 above) at 194.
Brownsword\textsuperscript{55} employs the Supreme Court of Israel decision in \textit{Jerusalem Community Burial Society} to illustrate this tension.\textsuperscript{56} Mr Kestenbaum entered into a contract with the burial society for his wife’s funeral arrangements. A term in the standard provided that the burial society would engrave their tombstones in characters from the Hebrew alphabet only. In accordance with his wife’s wishes, Mr Kestenbaum requested the burial society to engrave the tombstone with his wife’s name and Gregorian dates of birth and death in Latin characters, which the burial society refused by relying on the above term.\textsuperscript{57} The majority of the court annulled the contract term because the term, among other things, infringed upon Mr Kestenbaum’s freedom of expression, conscience and human dignity.\textsuperscript{58} In the dissenting judgment, it was held that respect for the free will of the contracting party is an essential public policy consideration and should be departed from in exceptional cases only. As the burial society’s decision to engrave all its tombstones with Hebrew characters promoted the dignity of the cemetery and took account of the feelings of the community members that used it, the dissenting court argued that the term did not negate public policy.\textsuperscript{59} Therefore, human dignity was used to argue both for the enforcement and annulment of the term which led Brownsword to argue that there is a tension between the two concepts.\textsuperscript{60}

This tension results from the fact that Brownsword and Beyleveld understand human dignity as empowerment to denote the freedom of a person to make her own choices without the interference of others. The moment her freedom (human dignity) is constrained to protect the freedom (human dignity) of another, there is a tension between the two persons’ freedom, and therefore, their human dignity. Defining human dignity as unconstrained freedom always results in one contracting party’s dignity being protected while the dignity of the other is compromised. As Brownsword articulates, the problem is then to explain why one contracting party ‘should be empowered at the expense’ of the other.\textsuperscript{61}

For years, the South African courts have grappled with this same question within, as well as outside, contract law – as will be illustrated below.\textsuperscript{62} Extrajudicially, one of the Justices of the appeal court even lamented that the constitutional value of human dignity ‘display[s] a perplexing capacity to pull in several directions at the same time’.\textsuperscript{63}

Next, I rely on Cornell and Wood’s communitarian interpretation of Kantian dignity to resolve this tension. I then argue that the Court is in the process of resolving this tension both outside and within the field of contract law by following this communitarian understanding of Kantian dignity.

\textsuperscript{55} Ibid at 182, 194–195.


\textsuperscript{57} \textit{Jerusalem Community} (note 56 above); Brownsword (note 25 above) at 182.

\textsuperscript{58} \textit{Jerusalem Community} (note 56 above). In his concurring judgment, Barak J held that the contract term violated the human dignity of the deceased and that of her family because neither she (during her lifetime) nor her family was allowed to determine the inscription on the tombstone. Ibid at para 28.

\textsuperscript{59} \textit{Jerusalem Community} (note 56 above) (Elon DP, dissenting judgment).

\textsuperscript{60} Brownsword (note 35 above) at 11.

\textsuperscript{61} Brownsword (note 25 above) at 194–195.

\textsuperscript{62} A discussion of the Court’s approach outside the field of contract law is discussed in part III and its approach within the realm of contracts is set out in part IV and V below.

\textsuperscript{63} Brand (note 19 above) at 86.
B  Reconciling the tension between human dignity as empowerment and constraint

In contrast to Brownsword and Beyleveld’s individualistic interpretation of Kantian dignity, which defines human dignity as empowerment as unconstrained freedom, Cornell and Wood propose a communitarian Kantian interpretation that defines freedom as a moral freedom which results in the reconciliation of the above tension.

Cornell argues that it is important to understand Kant’s conceptualisation of freedom and autonomy. According to Kant, freedom is a special type of causality which is a characteristic of human beings to the extent that they are rational. Kant distinguishes between ‘negative’ and ‘positive’ freedom. Negative freedom is a person’s ability to resist outside influences and her own immediate natural impulses and desires in order to act towards her own ends for her own long term wellbeing through practical reason. Kant defines freedom as a type of causality that operates according to a law and cannot be lawless, which leads him to argue that positive freedom refers to the idea that free will is the ability people to act in accordance with their own laws. Hence, positive freedom refers to the ability to make decisions in accordance with self-imposed laws.

Kant’s positive freedom should be understood in the context of one of his earlier formulations of the categorical imperative, the Formula of Autonomy, which refers to ‘the will of every rational being as a will which makes universal law’. For Kant, autonomy does not denote that human beings can do what they want without interference from others, but rather that they are bound only to those laws that they, as rational beings, can lay down upon themselves, provided such laws meet the standard of a universal law. Thus, a free will is a will under moral law, which means there is no tension between a person’s freedom and subjecting herself to the moral law. Therefore, Kantian freedom does not mean that you can do what you want without interference from others; it denotes the capacity to lay down moral laws for yourself to which you are bound. The constraint on your freedom is not external or in tension with your freedom as reflected in Beyleveld and Brownsword’s human dignity as constraint but is seen as a self-imposed constraint. Although Kantian dignity does not deny that human beings are autonomous beings who plot their own futures and make their own decisions, they...
must exercise their autonomy in a moral manner. Hence Kantian dignity does not support an understanding of human dignity as empowerment which denotes unconstrained freedom.\footnote{Beyleveld & Brownsword offer a differing interpretation of this passage. Beyleveld & Brownsword (note 29 above).}

The same type of reasoning applies when interpreting Kant’s second categorical imperative that requires that a person should ‘always treat humanity … in the person of any other, never simply as a means, but always at the same time as an end’.\footnote{Kant (note 27 above) at 91.} Wood emphasises that the idea of humanity as an end in itself denotes a person’s capacity to set ends for herself and choose ways to achieve them in order to lead a purposive life through her practical reason.\footnote{A Wood ‘Human Dignity, Right and the Realm of Ends’ (2008) \textit{Acta Juridica} 47, 53.} In other words, human dignity requires respect for the unique set of ends that an individual pursues. Accordingly, Wood arrives at the following understanding of what it entails to treat humanity in a person as an end in itself:

[H]uman beings should never be treated in a manner that degrades or humiliates them, should not be treated as inferior in status to others, or made subject to the arbitrary will of others, or be deprived of control over their own lives, or excluded from participation in the collective life of the human society to which they belong.\footnote{Ibid at 52.}

It is true that Wood’s interpretation of Kant’s second categorical imperative reflects human dignity as empowerment as respect for a person’s unique set of ends and her freedom to make choices without interference from others. However, this freedom is still a moral freedom that should not violate the human dignity of others. In other words, an individual’s choices should be respected provided such choices are not immoral.\footnote{S Woolman ‘Dignity’ in S Woolman & M Bishop (eds) \textit{Constitutional Law of South Africa} (2nd Ed, Revision Service 6, 2014) at para 36.2(c).} Therefore, when human dignity as empowerment is understood as denoting a moral freedom, it inevitably leads to the endorsement of the idea of human dignity as constraint as reflected in the correlating duty of an individual to respect the human dignity of others.

This leads the discussion to another formulation of Kant’s categorical imperative: the Formula of the Kingdom of Ends. Cornell and Wood continue to integrate Kant’s ‘kingdom of ends’ in their interpretation of human dignity, introducing the social ideals of Kantian ethics.\footnote{Wood (note 78 above) at 60; Cornell (note 65 above) at 30.} Kant’s hypothetical kingdom of ends provides that one should “[a]ct on the maxims of a member who makes universal laws for a merely possible kingdom of ends”.\footnote{Kant (note 45 above) at 100.} Kant defines his kingdom of ends as ‘a systematic union of different rational beings under common laws’.\footnote{Ibid at 95.} Wood explains the relevant passage\footnote{Ibid.} in Kant’s work:

A collection of \textit{ends} constitutes a ‘realm’ if these ends are not in conflict or competition with one another, but are combined into a mutually supporting system. … the \textit{Realm of Ends} commands us to follow maxims involving ends that belong to this system, and it forbids us to adopt ends that would stand in the way of rational beings sharing a system of ends. Ends that are neither required for nor incompatible with the system are permissible.\footnote{Wood (note 78 above) at 58.}
In a similar manner, Cornell argues that Kant’s hypothetical kingdom of ends is based on the idea that, as rational beings, we not only have the possibility of aligning our own actions with our own ends, but also with the ends of other rational beings. When we harmonise our own ends with the ends of others, we are aspiring to Kant’s imagined kingdom of ends. Kant explains that as members of the kingdom of ends, we are at the same time legislators and subjects of the moral law. Cornell emphasises that this capacity for self-legislation has a social dimension because we should attempt to harmonise our own ends with the ends of others. Therefore, in Kantian ethics, freedom always denotes a moral freedom and the categorical imperative requires human beings to harmonise their ends with the ends of others in a rational and moral manner.

Finally, it is important to note that Kant’s internal realm of freedom must be distinguished from the realm of external (legal) freedom which forms part of what Kant refers to as right (Recht). In Die Metaphysik der Sitten, Kant provides the following universal principle of right:

*Any action is right if it can coexist with everyone’s freedom in accordance with a universal law, or if on its maxim the freedom of choice of each can coexist with everyone’s freedom in accordance with a universal law.*

As emphasised by Wood, duties of right are normally enforced through state laws and therefore enforceable by coercion. This is in contrast with ethical duties which are an expression of the individual’s capacity to lay down self-imposed moral laws. Therefore, the exact relationship between the internal (moral) and external (legal) realm of freedom is uncertain and the subject of scholarly debate. An argument in favour of the idea that both are based on a unified principle is that Kant grounds the right to freedom on the humanity of human beings, and therefore, on the Formula of the End in Itself.

If it is accepted that the two realms are connected, how then is the internal (moral) realm of freedom represented in the external (legal) realm of freedom? Cornell answers as follows:

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.

In Cornell’s interpretation, the ideal of the kingdom of ends is utilised as a regulative ideal based on a moral social contract. When we harmonise our own ends with the ends of others

---

87 Ibid.
88 Kant (note 27 above) at 95.
89 Cornell (note 65 above) at 30.
90 Wood (note 78 above) at 54–55; Cornell (note 86 above) at 387.
91 Kant (note 45 above) at 24.
92 Wood (note 78 above) at 55.
93 Ibid at 56; Cornell (note 65 above) at 21. For a detailed discussion on the distinction between Kantian ethics and rights, see Wood (note 78 above) at 54–58.
94 Kant (note 46 above) at 30.
95 Cornell (note 86 above) at 387.
96 Cornell & Fuller (note 68 above) at 15. Although Wood argues that Kant’s realms of right and ethics are distinct systems with their own rational basis, he later argues that right can be used to further and promote the moral ideal of the kingdom of ends. Wood (note 78 above) at 57 and at 61.
as law-making members in the hypothetical kingdom of ends, we are in fact, reconciling our own freedom with the freedom of others.\footnote{Cornell & Fuller (note 68 above) at 15.} This is because Kantian freedom is internally self-limiting because freedom is equated to autonomy which involves laying down self-imposed moral laws.\footnote{Ibid at 14–15.} Therefore, when freedom is viewed in this way, it results in the resolution of the tension between human dignity as empowerment and human dignity as constraint because respecting the dignity of others is represented as a moral law we have imposed upon ourselves as law-making members in the imagined kingdom of ends.\footnote{Cornell (note 65 above) at 28.} Thus, when we disrespect the dignity of others we are actually failing to respect our own dignity as law-making members in the hypothetical kingdom of ends.\footnote{Cornell & Fuller (note 68 above) at 15.} We are not legislating together in a community that aspires to the kingdom of ends, and consequently, we are not free in the individual or collective sense.\footnote{Ibid.} As previously argued, the duty to respect the human dignity of others is a self-imposed constraint which we are able to exercise as part of our positive freedom, and ultimately, it is this capacity for positive freedom that gives us our dignity and our inherent worth.\footnote{Cornell (note 65 above) at 26.} Thus, when we disrespect the dignity of another, we are, in fact, disrespecting our own human dignity.\footnote{See the discussion in the text at note 75 above.}

C Why the communitarian interpretation of Kantian dignity is not enough: making the case for ubuntu

Although Cornell and Wood’s more communitarian interpretation of Kantian dignity provides a way to resolve the tension created by Beyleveld and Brownsword’s dualistic approach to human dignity, there are still a number of problems with Kantian dignity.\footnote{Cornell (note 86 above) at 388.} Cornell mentions four points of criticism against Kantian dignity to argue for a harmonisation between Kantian dignity and ubuntu.\footnote{Ibid.}

First, Kantian dignity is a Western conception of dignity which ignores indigenous communities’ understanding thereof and undermines proper legal pluralism as envisioned by the Constitution.\footnote{Cornell (note 65 above) at 19. See also Y Mokgoro & S Woolman ‘Where Dignity Ends and uBuntu Begins: An Amplification of, as well as an Identification of, a Tension in Drucilla Cornell’s Thoughts’ (2010) 25 South African Public Law 400, 402.}

Secondly, Kantian dignity does not take account of human beings who do not have the ability to act rationally and therefore Kantian dignity does not affirm the human dignity of all human beings.\footnote{Cornell (note 86 above) at 388.}

Thirdly, it has been argued that Kantian dignity is too individualistic. As shown above, such criticisms may ignore or misunderstand the more communitarian ideas in Kantian ethics. However, as acknowledged by Cornell, Kantian dignity is based on the individual’s capacity for reason. Therefore, Kant’s hypothetical moral social contract envisaged through the ideal of the kingdom of ends begins with imagined, already individuated individuals who, as free persons,
contract with each other for their own constraint in order to respect each other’s dignity.\textsuperscript{108} The moral social contract results in the maximisation of everyone’s unconstrained freedom in the external realm of right, which ultimately results in individual correlating rights and duties.\textsuperscript{109} This makes it difficult to use the Kantian imagined social contract to justify socio-economic rights and the promotion of substantive equality\textsuperscript{110} because both require greater constraint on the freedom of certain members of the community in order to ensure and promote the social and economic equality of all community members.

Fourthly, substantive equality requires an understanding of the existing social and economic differences within a society which involves taking account of the actual lived social and economic conditions of the community members. This is different from Kantian dignity which does not take account of the actual lived circumstances of human beings because it is based on an ideal\textsuperscript{111} and human beings’ capacity for practical reason whether they actually use this capacity or not.\textsuperscript{112} Although Wood concedes that Kantian dignity does not automatically implicate measures to address social inequality, he argues that it can be used as a critique against the ideologies underlying the free market economy which portrays human dignity as unconstrained freedom while in reality the weaker contracting party is in an unequal bargaining relationship and cannot be regarded as free or equal.\textsuperscript{113}

D Human dignity through ubuntu

The premise of this article is that the above criticisms can be addressed by ubuntu. To understand how this harmonisation can be achieved, it is necessary to discuss how human dignity is understood through ubuntu. It is not my intention to make an original contribution to the understanding of ubuntu in African philosophy as its meaning is the subject of some controversy which has resulted in criticism against its use as a legal concept.\textsuperscript{114} The best approach to establish the legal meaning of ubuntu is to focus on the judicial descriptions of ubuntu.\textsuperscript{115} This article draws on the work of Cornell including writings she produced in collaboration with Muvangua relying on prominent scholars in African philosophy and African jurisprudence to elucidate the role and meaning of ubuntu as found in South African law in order to focus attention on the distinction between Kantian dignity and Ubuntu.\textsuperscript{116} For the purposes of this article, a brief summary of their arguments is presented here.\textsuperscript{117}

\footnotesize

\begin{itemize}
\item \footnotesize 109 Ibid at 211.
\item \footnotesize 110 Ibid.
\item \footnotesize 111 Cornell (note 74 above) at 667.
\item \footnotesize 112 Cornell (note 86 above) at 388.
\item \footnotesize 113 Wood (note 78 above) at 61–64.
\item \footnotesize 114 I Keevy “The Constitutional Court and Ubuntu’s “Inseparable Trinity”” (2009) 34 Journal for Juridical Science 61(Ubuntu as a religious worldview that violates s 15(1) of the Constitution); I Keevy ‘Ubuntu Versus the Core Values of the South African Constitution’ (2009) 34 Journal of Juridical Science 19–58 (Ubuntu as a part of African law and African thinking that is patriarchal in nature and against the foundational constitutional values).
\item \footnotesize 115 C Himonga ‘Exploring the Concept of Ubuntu in the South African Legal System’ in U Kischel & C Kirchner (eds) Ideologie und Weltanschauung im Recht (2012) 1, 2.
\item \footnotesize 117 See Du Plessis (note 2 above) at 225–267 (Detailed discussion of their arguments and the African sources they rely on).
\end{itemize}
Ubuntu has been linked to the moral theory of humanism and the political and economic theory of socialism. These associations are also noticeable in judicial descriptions of ubuntu. In *Makwanyane*, it was held that ubuntu’s ‘spirit emphasises respect for human dignity’, and since then, the Court has reiterated this link between ubuntu and human dignity on a number of occasions. In *Makwanyane*, ubuntu was also linked to the ideal of social justice and in a number of cases thereafter it was used to promote the realisation of socio-economic rights.

While it can be deduced that ubuntu emphasises respect for human dignity and the promotion of substantive equality, it is necessary to understand how these ideas are understood in African thinking. In *Makwanyane*, Mokgoro J defined ubuntu as ‘humaneness’ which she translated as ‘personhood’ and ‘morality’. To understand what she meant by these terms, it is necessary to understand how ubuntu denotes the moral development of the individual and the African social bond. Cornell relies on Menkiti to conclude that a human being is intertwined in ethical relations with other community members and that her humaneness is embedded in the community from birth. Therefore, this social bond is a social fact and not an imagined social contract as found in Kantian dignity. Furthermore, it is only through actual ethical relations and engagement with community members and their support that a person can and must develop towards becoming an individuated moral being. However, this social bond should not be seen as simple communitarianism which denies individual interests. Cornell and Muvangua propose that the reason why Mokgoro J linked the term ‘humaneness’ to personhood and morality was to denote how a person’s development into a unique being (her personhood) is inseparable from her moral development. Therefore, like the communitarian interpretation of Kantian dignity, ubuntu emphasises the inseparable link between morality and freedom. Hence both Kantian dignity and ubuntu understand human dignity as empowerment as denoting a moral freedom. However, as ubuntu reflects

119 More (note 118 above) at 156; Mokgoro (note 118 above) at 3.
122 *Makwanyane* (note 120 above) at para 237 (Madala J).
123 *Port Elizabeth* (note 121 above).
124 *Makwanyane* (note 120 above) at para 308.
125 Cornell (note 86 above) at 392.
127 Cornell (note 86 above) at 392.
128 Ibid.
131 Cornell & Muvangua (note 116 above) at 8.
132 Cornell (note 86 above) at 397.
that a person’s dignity is rooted in her personhood (uniqueness) and her embeddedness in the community, it differs from Kantian dignity which is based on the capacity for rationality. Therefore, unlike Kantian dignity, ubuntu confirms the human dignity of every person whether she has the capacity to act rationally or not.

As the community must respect and support the moral development of the individual into a unique being, the community has a duty to respect the human dignity of the individual. In the same way, the individual has a duty to respect the human dignity of other community members. Therefore, ubuntu accords with the idea of human dignity as constraint as reflected in the correlating duty to respect the human dignity of others as based on Kantian dignity. However, this duty goes further than Kantian dignity because it is also concerned with the actual wellbeing of community members. Cornell explains that each community member has a duty to support and promote the community that supports her. Masolo speaks of the duty each person has to share in the responsibility of bringing about an ethical and more humane world which results in ‘a thick system of rights and obligations’. This duty entails more than an individual duty that correlates with a specific individual right, and goes further than the harmonisation of individual freedoms to maximise the unconstrained freedom of everyone as reflected in the Kantian right to freedom. Ubuntu is also concerned with the realisation and promotion of the socio-economic wellbeing of all the community members.

Cornell further contends that ubuntu ‘has an aspirational and ideal edge’ because bringing about a humane world and becoming a moral person in that world is a never-ending task. However, unlike the Kantian ideal of the kingdom of ends that is a regulative ideal, ubuntu is embedded in the social reality and materialises in the ethical actions between community members. It is especially concerned with the wellbeing and welfare of community members because it recognises that the identity and dignity of a human being is embedded in the community. Hence, it takes account of the actual social and economic circumstances of community members. Therefore, as Cornell concludes, ubuntu is a moral demand for social justice and harmony, and consequently promotes the realisation of socio-economic rights and substantive equality. In this way, ubuntu can be linked to the idea of human dignity.

133 Ibid at 397.
134 Makwanyane (note 120 above) at para 309 (Mokgoro J links the inherent dignity of every person to Ubuntu).
135 Ibid at para 224 (Langa J).
136 Compare the discussion on human dignity as constraint in the text at note 44 above.
137 Cornell (note 86 above) at 393. See also D Cornell & K van Marle ‘Ubuntu Feminism: Tentative Reflections’ (2015) 36(2) Verbum et Ecclesia 1, 2; Cornell & Muvangua (note 116 above) at 4.
139 Cornell (note 86 above) at 393.
140 Compare the discussion on Kant’s right to freedom in the text at note 91 above and the criticism levelled against Kantian dignity as discussed in the text at note 108 above.
141 Bhe & Others v Magistrate, Khayelitsha, & Others (Commission for Gender Equality as Amicus Curiae); Shibi v Sithole & Others; South African Human Rights Commission & Another v President of the Republic of South Africa & Another [2004] ZACC 17, 2005 (1) SA 580 (CC)(‘Bhe’) at para 163. See also Makwanyane (note 120 above) at para 224.
142 Cornell (note 86 above) at 396. See also Makwanyane (note 120 above) at para 227 (Langa J).
143 Cornell (note 86 above) at 396.
144 Compare the discussion on Kant’s right to freedom in the text at note 118 above at 157.
145 Port Elizabeth (note 121 above) at para 31(Sachs J insists that the particular circumstances of each case must be considered.)
146 Cornell (note 86 above) at 396–397.
as constraint as reflected in the duty on the individual to realise the human dignity of other persons through the promotion of socio-economic rights and substantive equality.\textsuperscript{147} This is in contrast to Kantian dignity which struggles to support the justification and promotion of socio-economic rights and substantive equality.

Cornell and Muvanguard argue that as a person’s identity, and attending human dignity, is embedded in the community, the community members are, in a sense, a part of her, because her moral development into personhood happens through her engagement with other community members and her participation in the community’s practices and traditions.\textsuperscript{148} This idea that the wellbeing and human dignity of the individual is tied up with the wellbeing and human dignity of the other community members is sometimes referred to as human dignity as a collective responsibility or concern and has been utilised by the Court to promote the realisation of socio-economic rights.\textsuperscript{149} Although Justice Mokgoro did not explicitly refer to ubuntu in \textit{Khosa}, dealing with the state’s refusal to provide social welfare benefits to permanent residents, it has been argued\textsuperscript{150} that her following remark is based on ubuntu:

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.\textsuperscript{151}

## E Harmonisation of Kantian dignity and ubuntu

In view of the above analysis, Cornell identifies two similarities between Kantian dignity and ubuntu.\textsuperscript{152} Both emphasise the link between freedom and morality. Both also promote human dignity in that they emphasise the inseparableness of personhood and morality.\textsuperscript{153} Hence both the communitarian interpretation of Kantian dignity and ubuntu view empowerment conception of human dignity as denoting a moral freedom which necessarily results in the endorsement of human dignity as constraint reflected in the duty to respect the human dignity of others.

Cornell also identifies two important differences between Kantian dignity and ubuntu. First, where Kantian dignity is based on the rational capacity of human beings, dignity through ubuntu is based on the uniqueness of human beings and their embeddedness in the community. Therefore, unlike Kantian dignity, ubuntu recognises the inherent dignity of all human beings.\textsuperscript{154} Secondly, the social bond is construed differently in African thinking than that envisaged by Kantian dignity.\textsuperscript{155} Kantian dignity as based on a hypothetical social contract results in the maximisation of everyone’s unconstrained freedom. In contrast, ubuntu relies on

---

\textsuperscript{147} Compare the discussion dealing with human dignity as constraint in the text at note 48 above.

\textsuperscript{148} Cornell (note 86 above) at 393. See also Pillay (note 121 above) at para 53.

\textsuperscript{149} Albertyn (note 18 above) at 344 and Mokgoro & Woolman (note 106 above) at 403, footnote 10.

\textsuperscript{150} Albertyn (note 18 above; Cornell & Muvanguard (note 116 above) at 20; Mokgoro & Woolman (note 106) at 403; Cornell &Van Marle (note 108 above) at 213.

\textsuperscript{151} \textit{Khosa & Others v The Minister of Social Development & Others; Mabulae & Others v Minister of Social Development & Others} [2004] ZACC 11, 2004 (6) SA 505 (CC) (‘\textit{Khosa}’) para 74. See also \textit{Port Elizabeth} (note 121 above) at para 18 (Sachs J explicitly relied on ubuntu to make a similar point).

\textsuperscript{152} Cornell (note 86 above) at 397.

\textsuperscript{153} Ibid at 398.

\textsuperscript{154} Ibid at 398–399.

\textsuperscript{155} Ibid at 399.
a social bond that is a social fact which takes account of the lived circumstances of community members and entails a duty on the individual to recognise and support the socio-economic wellbeing of other community members.^{156} Thus, while both Kantian dignity and ubuntu support the idea of human dignity as constraint as reflected in the individual’s correlating duty to respect the human dignity of others, ubuntu goes further in that it also denotes an additional duty on the individual to assist in the realisation and promotion of socio-economic rights and substantive equality.

For Wood, the Kantian ideal of the kingdom of ends resonates with the concept of ubuntu, specifically the idea that a human person becomes a person through other human beings.\textsuperscript{157} Wood argues that relating Kant’s ideal kingdom of ends to ubuntu can assist in ‘interpreting the South African Constitution’s commitment to the inherent dignity of every human being’.\textsuperscript{158} It is perhaps in this light that Wood’s proposal, that Kant’s kingdom of ends could be used to critique the individualistic notion of human dignity which informs the ideologies of individualism and free trade, can be seen.\textsuperscript{159} Hence, it can be argued that it is through ubuntu that the Kantian ideal of the kingdom of ends can be developed to take account of and promote socio-economic rights and substantive equality. As will be seen below, the Court also relied on Kant’s kingdom of ends in promoting socio-economic rights and substantive equality within a larger understanding of human dignity as informed by ubuntu.\textsuperscript{160}

III THE COURT’S APPROACH TO HUMAN DIGNITY

The Court in \textit{Makwanyane} confirmed that human dignity refers to the intrinsic worth of all human beings which means that they are entitled to equal respect and concern and may not be treated in a degrading or dehumanising way.\textsuperscript{161} The Kantian idea that dignity is beyond price and of incalculable worth, as well as Kant’s second categorical imperative (treating humanity as an end in itself), was explicitly imported into the constitutional jurisprudence by Ackermann J in \textit{Dodo}.\textsuperscript{162} In \textit{Khumalo}, O’Regan J confirmed that human dignity does not only refer to a person’s own self-worth but also how she is valued and treated by the members of the community.\textsuperscript{163} Therefore, the Court has confirmed the inherent and equal worth of all human beings that results in their entitlement to equal respect and concern as based on Kantian dignity.\textsuperscript{164} As the Court has also relied on ubuntu to confirm the inherent dignity of every person,\textsuperscript{165} and because it is better suited to the task,\textsuperscript{166} this can be viewed as a harmonisation between Kantian dignity and ubuntu.

\begin{footnotes}
\item[156] Cornell & Van Marle (note 137 above) at 3.
\item[157] Wood (note 78 above) at 60.
\item[158] Ibid at 60–61.
\item[159] Compare the discussion in the text at note 113 above.
\item[160] Compare the discussion in the text at note 181 below.
\item[161] \textit{Makwanyane} (note 120 above) at paras 26 (Chaskalson J), at paras 271 and 281 (Mohamed J), at paras 313–316 (Mokgoro J) and at para 328 (O’Regan J).
\item[162] \textit{S v Dodo} [2001] ZACC 16, 2001 (3) SA 382 (CC) (‘\textit{Dodo}’) at para 38. See also Albertyn (note 18 above) at 327, fn 28; Botha (note 36 above) at 202.
\item[163] \textit{Khumalo & Others v Holomisa} [2002] ZACC 12, 2002 (5) SA 401 (CC) (‘\textit{Khumalo}’) at para 27
\item[165] Compare the discussion in the text at note 120 above.
\item[166] Compare the discussion in the text at note 154 above.
\end{footnotes}
In *Pillay*, dealing with a school learner’s right to wear a gold nose-stud in accordance with her South Indian family traditions and culture, the Court explained what it means to treat human beings as an end in themselves. The Court emphasised the importance of freedom in defining human dignity\(^{167}\) by quoting from Ackermann J’s minority judgment in *Ferreira* where he stated that ‘[h]uman dignity has little value without freedom; for without freedom personal development and fulfilment are not possible’.\(^{168}\) Langa J then continued as follows:

A necessary element of freedom and of dignity of any individual is an ’entitlement to respect for the unique set of ends that the individual pursues’. One of those ends is the voluntary religious and cultural practices in which we participate.\(^{169}\)

The Court’s conception of what it means to treat a human being as an end in herself reflects Wood’s interpretation of Kant’s second categorical imperative. As was shown above,\(^{170}\) Wood proposed that treating a person’s humanity as an end in itself entails respecting her capacity to set her own ends and choose ways to achieve them in order to lead a purposive life through her practical reason. For Wood, like Beyleveld and Brownsword, this means that the autonomy of a person must be respected by not subjecting another’s arbitrary will onto that person. Furthermore, Wood emphasised that this means that a person should not be excluded from participating in community life and would include respect for a person’s participation in religious or cultural practices.\(^{171}\) In similar fashion, Botha interpreted the Court’s finding to mean that respecting a person as an end in herself ‘demands the creation of a space within which individuals are free to forge their own autonomous identities’.\(^{172}\) Therefore, the Court has developed the constitutional value of human dignity to reflect its empowerment conception which denotes the right to a space in which a person’s human dignity can flourish. In this respect, the court also confirmed the state’s duty to create and protect such a space for individual autonomy as reflected in the notion of human dignity as constraint.

Although the Court has emphasised the link between human dignity and autonomy in a number of cases,\(^{173}\) this does not mean that the Court views this freedom as an unconstrained freedom or the individual ‘as an isolated and unencumbered being’.\(^{174}\) In dealing with the right to privacy in *Bernstein*, Ackermann J emphasised the correlating nature of rights and duties:

> The truism that no right is to be considered absolute implies that from the outset of interpretation each right is always already limited by every other right accruing to another citizen. … This implies that community rights and the rights of fellow members place a corresponding obligation on a citizen.\(^{175}\)

It would seem that Ackermann J was influenced by Kant’s imagined kingdom of ends as he refers to Kant’s ‘community of humanity’ which ‘demands mutual respect as a universal

---

167 *Pillay* (note 121 above) at para 63.
168 *Ferreira v Levin NO & Others; Vryenboek & Others v Powell NO & Others* [1995] ZACC 13, 1996 (1) SA 984 (CC) (‘*Ferreira*’) at para 49.
169 *Pillay* (note 121 above) at para 64.
170 Compare the discussion in the text at note 78 above.
171 Compare the discussion in the text at note 79 above.
172 Botha (note 36 above) at 203–204.
173 *MM v MN & Another* [2013] ZACC 14, 2013 (4) SA 415 (CC) at paras 73–74; *Teddy Bear Clinic* (note 16 above) at para 56; *Barnard* (note 164 above) at para 173 (Van der Westhuizen J).
174 *Barnard* (note 164 above) at para 174. See also *Bernstein & Others v Bester and Others NNO* [1996] ZACC 2, 1996 (2) SA 751 (CC) (‘*Bernstein*’) at para 65; Ackermann (note 121 above) at 109–111.
175 *Bernstein* (note 174) at para 67.
moral duty towards persons as *moral persons*.\(^\text{176}\) This can be viewed as an endorsement of the communitarian conception of human dignity as empowerment which entails a moral freedom rather than an unconstrained freedom. As acknowledged by Ackermann J, such an understanding of human dignity necessarily implicates the correlating duty of an individual to respect the human dignity of others as reflected in the idea of human dignity as constraint.

The Court has also relied on ubuntu to emphasise that ‘human beings are social beings whose humanity is expressed through their relationships with others’.\(^\text{177}\) In *Pillay* the Court relied on ubuntu to emphasise that a person’s moral development into a unique human being can only happen through engagement with other community members, and therefore, participation in the community’s practices and traditions is inseparably linked to a person’s human dignity.\(^\text{178}\) As the Court also referred to Kant’s second categorical imperative to promote the idea that human dignity denotes respect for a person’s unique set of ends, this case illustrates how both Kantian dignity and ubuntu promote human dignity in a way that protects a person’s individuality and provides a space for that person to pursue her moral personhood and unique destiny (ubuntu); or her unique moral set of ends (Kantian dignity) within and through engagement with the community which can only happen when the community members respect the human dignity of an individual by creating a space for her to pursue her unique destiny or ends.\(^\text{179}\) In this way, the Court has harmonised Kantian dignity and ubuntu to endorse the empowerment conception of human dignity as a moral freedom which necessarily implicates the correlating duty of an individual to respect the human dignity of others as reflected in human dignity as constraint.

Nevertheless, this is not the end of the matter because it is difficult to use Kantian dignity to promote the realisation of socio-economic rights and substantive equality in the private sphere. As explained, human dignity based solely on Kantian dignity does not necessarily implicate a duty on the individual to promote social justice in private dealings.\(^\text{180}\) The solution to this problem lies with ubuntu as it is more suited to the task\(^\text{181}\) and it has proved particularly valuable in infusing the constitutional value of human dignity with the duty on the state to promote the realisation of socio-economic rights and substantive equality.\(^\text{182}\) Recently, in *Barnard*, Van der Westhuizen J extended the application of these duties to individuals when dealing with promotion of substantive equality within the workplace. He relied on both ubuntu\(^\text{183}\) and Kantian dignity\(^\text{184}\) to make the following statement:

\(^\text{176}\) Ibid at para 66, fn 93.

\(^\text{177}\) *Dawood & Another v Minister of Home Affairs & Others; Shalabi & Another v Minister of Home Affairs & Others; Thomas & Another v Minister of Home Affairs & Others* [2000] ZACC 8, 2000 (3) SA 936 (CC) (*Dawood*) para 30, fn 42.

\(^\text{178}\) *Pillay* (note 121 above) at para 53.

\(^\text{179}\) Botha (note 36 above) at 205.

\(^\text{180}\) Compare the discussion in the text at note 108 above.

\(^\text{181}\) Compare the discussion in the text at note 155 above.

\(^\text{182}\) Compare the discussion in the text at note 123 above. See also Himonga (note 115 above) at 15, fn 61.

\(^\text{183}\) Van der Westhuizen J referred to the following decisions in which ubuntu was used to develop the constitutional value of human dignity: *Pillay* (note 121 above); *Port Elizabeth* (note 121 above); *Dawood* (note 177 above).

\(^\text{184}\) Van der Westhuizen J referred to the following decisions that incorporate aspects of Kantian dignity into South African law: *Khumalo* (note 163 above); *Dodo* (note 162 above); *Ferreira* (note 168 above); *Bernstein* (note 174 above); *Makwanyane* (note 120 above); *National Coalition for Gay and Lesbian Equality & Another v Minister of Justice & Others* [1998] ZACC 15, 1999 (1) SA 6 (CC).
In the context of socio-economic rights, this Court has affirmed that the responsibility for the difficulties of poverty is shared equally as a community because ‘wealthier members of the community view the minimal well-being of the poor as connected with their well-being and the well-being of the community as a whole’. This would also hold in the context of substantive equality. First, the way in which individuals interact with social groups and society generally has a direct bearing on their dignity. This is true for members of both advantaged and disadvantaged groups. Second, this idea also gives effect to another Kantian way of understanding dignity – that it ‘asks us to lay down for ourselves a law that embraces every other individual in a manner that extends beyond the interests of our more parochial selves’.  

In the first place, Van der Westhuizen J’s statement supports the premise that the Court is following the more communitarian interpretation of Kantian dignity that reconciles the tension between human dignity as empowerment and constraint. Cornell argued that Kantian dignity denotes the capacity to lay down moral laws for ourselves and being bound by those laws. This means that freedom can be found only in the realm of morality and there is no tension between our freedom and subjecting ourselves to the moral law. Relying on Kant’s ideal kingdom of ends, she concluded that when we disrespect the dignity of others we are actually failing to respect our own dignity as law-making members in the kingdom of ends. These ideas are reflected in Van der Westhuizen J’s reasoning when he maintains that substantive equality measures ‘can enhance the dignity of individuals, even those who may be adversely affected by them’. Secondly, as was shown above, ubuntu also recognises that the social and economic wellbeing and human dignity of the individual is tied up with the social and economic wellbeing and human dignity of the other community members. This is also reflected in Van der Westhuizen J’s reasoning where he states that ‘wealthier members of the community view the minimal well-being of the poor as connected with their well-being and the well-being of the community as a whole’. Therefore, the Court is in the process of developing the constitutional value of human dignity to include a duty on the individual to assist in the promotion of other community members’ socio-economic wellbeing and the realisation of substantive equality, and it is drawing from both Kantian dignity and ubuntu to do so.  

To summarise, the Court is in the process of developing a communitarian understanding of human dignity through the harmonisation of Kantian dignity and ubuntu. Both Kantian dignity and ubuntu have been used to endorse the idea of human dignity as empowerment as a moral freedom which results in an individual’s correlating self-imposed duty to respect the human dignity of others. Furthermore, this conceptualisation of human dignity as constraint denotes not only a duty to respect the human dignity of others (a harmonisation between Kantian dignity and ubuntu) but also a duty to share in the responsibility of creating the necessary conditions for the realisation of others’ human dignity through the promotion of socio-economic rights and substantive equality (a development inspired by ubuntu).  

In order to apply this sophisticated understanding of human dignity in the common law of contract, I revisit the three major cases dealing with contractual justice (Barkhuizen, Bredenkamp and Botha) and propose a new interpretation of Botha that supports the creation of an equitable discretion to set aside unfair contract terms as well as the unfair enforcement thereof.

---

185 Barnard (note 164 above) at para 175 quoting from Khosa (note 151 above) at para 74.  
186 Compare the discussion of Cornell’s interpretation of Kantian dignity in the text at note 65 above.  
187 Barnard (note 164 above) at para 175.  
188 Compare the discussion in the text at note 148 above.
IV NECESSARY CONTEXT: CONTRACTUAL FAIRNESS PRE- AND POST CONSTITUTION

Before the Court’s sophisticated understanding of human dignity can be applied to these three cases, it is necessary to place them in the context of the legal position in respect of unfair contract terms immediately prior to the enactment of the Constitution as well as the most relevant decisions handed down thereafter.

A Contractual fairness prior to the Constitution: Sasfin

Sasfin is the foremost decision dealing with contractual fairness prior to the Constitution. An anaesthetist (Beukes) granted a deed of cession in favour of a finance company (Sasfin) which placed Sasfin in complete control of his earnings. In terms of the deed, on notice of cession to Beukes’ debtors, Sasfin would be able to recover all of Beukes’ book debts and retain all amounts recovered, whether or not he owed any money to Sasfin. Beukes was incapable of ending this situation.

The court held that a contractual term which is contrary to public policy is unenforceable. It accepted that public policy is a vague and contentious concept, in other words an open norm, but argued that:

[...]gements which are clearly inimical to the interests of the community, whether they are contrary to law or morality, or run counter to social or economic expediency, will accordingly, on the grounds of public policy, not be enforced.

Therefore, a court must not hesitate to declare a contract contrary to public policy when necessary, but such power should be exercised with restraint to ensure that the decision does not cause legal uncertainty and is not based on the judge’s own subjective perception of fairness.

The court stressed that, on the one hand, public policy generally favours freedom and sanctity of contract. On the other hand, it should also take account of ‘the doing of simple justice between man and man’. Finally, in balancing these policy considerations, the court concluded that the provisions of the contract were unconscionable and therefore contrary to public policy. Specifically, the court held that the contract relegated Beukes to the position of a slave in that he was working for the benefit of Sasfin, could be deprived of all his income and was unable to end this situation.

Although the court did not refer to human dignity expressly, it is possible to fit the court’s decision within the matrix of the communitarian understanding of human dignity as based on Kantian dignity. In the first place, the public policy consideration of freedom and sanctity of contract is an expression of the idea of human dignity as empowerment in that it protects and promotes a contracting party’s contractual autonomy in concluding a contract. Therefore, this public policy consideration supports the enforcement of a contract. However, the court does not view this freedom as an unconstrained freedom that may not be interfered with. Contracts that are ‘inimical to the interests of the community’, whether illegal, immoral or

---

189 Sasfin (Pty) Ltd v Beukes 1989 (1) SA 1 (A) (‘Sasfin’) 7.
190 Ibid at 7, 13–14.
191 Sasfin (note 189 above) at 7–8.
192 Ibid.
193 Ibid at 9.
194 Ibid at 13–14.
195 Compare discussion in the text at note 40 above.
socially and economically inexpedient, should not be enforced. These references align with the Kantian ideal kingdom of ends which is based on a social contract in terms of which each community member, as a free individual, has contracted to respect the human dignity of others. Hence a party’s contractual autonomy is a moral freedom that may not be exercised in a way that would infringe upon the human dignity of the other party. As shown above, this approach necessarily leads to the endorsement of the idea of human dignity as constraint as reflected in the duty to respect the human dignity of others which can be used to justify the non-enforcement of a contract. In can be argued that the Sasfin court expressed this duty through the public policy consideration of simple justice between man and man.

Relying on Kant’s second categorical imperative (treating the humanity in another as an end in itself), it can be argued that as the terms of the contract deprived Beukes of control over his life to such an extent that it relegated him to the position of a slave, the contract allowed Sasfin to treat Beukes merely as a means to an end and not as an end in himself, and accordingly, the contract infringed upon Beukes’ human dignity. For this reason, the court concluded that the contract was unconscionable and therefore against public policy.

It is important to note that the term ‘unconscionable’ is an open-ended term and therefore open to different interpretations which enable value judgements. So how to explain the court’s resolution of the tension between human dignity as empowerment and constraint through a value judgment that necessarily entails a consideration of the fairness of the contract terms? Lubbe argued that Kant’s second categorical imperative does not mean that a contracting party should never treat the other contracting party as a means to her own ends, but rather that she must never treat the other contracting party merely as a means, but always also as an end. He argued that Kant’s second categorical imperative ‘suggests that in the pursuit of one’s own ends, a minimum level of regard for the interests of others is indicated’. This is because contracts are typically bilateral co-operative ventures in terms of which both parties have to perform and both benefit from the contractual relationship in some manner. In other words, each party’s contractual performance contributes to the other party’s pursuit of her unique ends, while, at the same time, each party also uses the other party as a means in pursuit of her own ends. Hence, contractual terms in a bilateral contract will always reflect two unique set of ends that must be harmonised with each other as reflected in Kant’s ideal kingdom of ends. Ultimately, this means that when determining whether a contractual term in such a contract infringes the human dignity of one of the parties, a court has to balance the interests of both parties (as reflected in the contract terms themselves) in order to determine whether

---

196 Compare Cornell’s interpretation of Kant’s freedom as a moral freedom (discussion in the text at note 65 above).
197 Compare the discussion in the text at note 80 above.
198 Compare the discussion in the text at note 52 above.
200 The court accepted that determining whether a contract is against public policy involves a value judgment. Compare the discussion in the text at note 191 above.
201 Lubbe (note 19 above) at 421, fn 188 referring to Beyleveld & Brownsword (note 29 above) at 666, footnote 30. See also Woolman (note 80 above) at para 36.2(b).
202 Lubbe (note 19 above) at 421, fn 188.
203 G Lubbe ‘Bona fides, Billikheid en die Openbare Belang in die Suid-Afrikaanse Kontraktereg’ (1990) 1 Stellenbosch Law Review 7, 20 (Speaks of the need to harmonise conflicting individual interests). Compare Lubbe’s views with Cornell’s interpretation of Kant’s kingdom of ends which entails the harmonisation of ends as discussed in the text at note 86 above.
the term constitutes an unfair infringement upon one of the party’s interests (her unique set of ends) and hence violates her human dignity.

By following an approach that aligns with the communitarian understanding of Kantian dignity, the appeal court endorsed the idea of human dignity both as empowerment which denotes a moral freedom and as a constraint reflected in the duty on the individual to respect the human dignity of others. According to the communitarian understanding of Kantian dignity, this constraint is not in tension with the contracting party’s autonomy (in other words the party who is prevented from enforcing the contract) because respecting the human dignity of the other party is seen as a self-imposed moral duty that the constraining party has laid down for herself in terms of a social contract.

Therefore, determining whether a contract term is contrary to public policy involves a balancing act between competing values\(^{204}\) and the balancing act will always be between policy considerations that support the enforcement of the contract, on the one hand (as a reflection of human dignity as empowerment), and policy considerations that support the non-enforcement of the contract, on the other (as reflected in human dignity as constraint). Ultimately, this balancing act involves a value judgment to determine the fairness of the terms. In light of this interpretation, the decision in *Sasfin* can be used to support the existence of the court’s equitable discretion to set aside unfair contract terms on the basis of human dignity although the court did not expressly rely on human dignity to reach its decision.

However, the court had a very limited understanding of fairness as it confined itself to an investigation of the fairness of the contractual terms themselves. The court did not take account of the circumstances surrounding the conclusion of the contract or those at the time of its enforcement. In other words, the court did not take account of the socio-economic reality of the contracting parties or the idea of substantive equality. Therefore, although the appeal court’s understanding of contractual fairness reflects an understanding of Kantian dignity that entails a moral freedom and the resulting duty to respect the human dignity of others, it does not require an individual to assist in the promotion of the socio-economic wellbeing of the other community members in her private dealings with them. Thus, at the most, *Sasfin* endorses the communitarian understanding of Kantian dignity which results in the maximisation of everyone’s unconstrained freedom as reflected in Kant’s external realm of right and this is an expression of a more liberal understanding of human dignity that does not promote socio-economic rights or substantive equality.\(^{205}\)

The appeal court did not refer to the concept of good faith (another open norm), presumably because there was some uncertainty regarding its specific role in promoting contractual fairness in the aftermath of *Bank of Lisbon*.\(^{206}\) However, Lubbe argued that the principle of good faith could be developed to perform the same function as the policy consideration of simple justice between man and man and that where a contractual term constitutes an unreasonable and one-sided promotion of one party’s own interest at the expense of the other party it may be

---


\(^{205}\) Compare the discussion in the text at note 107 above with the line of criticism lodged against Kantian dignity as too individualistic because it does not promote socio-economic rights or substantive equality.

\(^{206}\) *Bank of Lisbon and South Africa Ltd v De Ornelas* [1988] ZASCA 35, 1988 (3) SA 580 (A)(‘Bank of Lisbon’) (Appeal court eliminated the *exceptio doli* and did not transfer its role to the concept of good faith, but held that all contracts are *bonae fidei*. This resulted in some uncertainty in re the exact role of good faith in preventing and correcting contractual unfairness). See also Du Plessis (note 2 above) at 135–141.
contrary to good faith and consequently also against public policy.\textsuperscript{207} Similarly, Barnard argued that good faith can be seen as an expression of the duty of an individual to respect the human dignity of others (as reflected in human dignity as constraint) when applied to the common law of contract.\textsuperscript{208}

Lubbe’s approach was followed by appeal judge Olivier in his minority judgment in \textit{Saayman}.\textsuperscript{209} He argued that the \textit{Bank of Lisbon} court did not limit the role of good faith in the common law of contract.\textsuperscript{210} He maintained that a court may apply the principles of public interest which include the principle of good faith, but conceded that this must be done carefully to prevent legal uncertainty and the arbitrary decisions.\textsuperscript{211} He further stated that public interest does not demand that a party should be held bound to an otherwise enforceable contract where such a contract falls foul of the requirements of good faith.\textsuperscript{212} Soon thereafter, Olivier J’s approach was followed by the Cape division in \textit{Miller},\textsuperscript{213} and later that same year, the appeal court also referred to his judgment with apparent approval.\textsuperscript{214}

B Contractual fairness after the Constitution

Initially, there were indications that the multi-faceted idea of human dignity as empowerment (denoting a moral freedom) and constraint (as reflected in the resultant duty to respect the human dignity of others) – the latter finding expression through good faith – would find application in the common law of contract. In \textit{Mort}, Davis J stated \textit{obiter} that the court must develop the common law of contract in line with the Constitution.\textsuperscript{215} He maintained that good faith is shaped by the community’s legal convictions with reference to the constitutional values of human dignity, equality and freedom.\textsuperscript{216} Freedom supports freedom and sanctity of contract while equality and dignity support the idea that contracting parties –

\begin{quote}
must adhere to a minimum threshold of mutual respect in which ‘the unreasonable and one-sided promotion of one’s own interest at the expense of the other infringes the principle of good faith to such a degree as to outweigh the public interest in the sanctity of contracts’.\textsuperscript{217}
\end{quote}

\begin{thebibliography}{99}
\item \textsuperscript{207} Lubbe (note 203 above) at 17–21.
\item \textsuperscript{208} Barnard (note 74 above) at 289.
\item \textsuperscript{209} \textit{Eerste Nasionale Bank van Suidelike Afrika Bpk v Saayman NO} [1997] ZASCA 62, 1997 (4) SA 302 (SCA) (‘\textit{Saayman}’) at 318.
\item \textsuperscript{210} Ibid at 323.
\item \textsuperscript{211} Ibid at 324 referring to \textit{Sasfin} (note 189 above) at 9.
\item \textsuperscript{212} \textit{Saayman} (note 209 above) at 331.
\item \textsuperscript{214} \textit{NBS Boland Bank Ltd v One Berg River Drive CC & Others; Deeb & Another v ABSA Bank Ltd; Friedman v Standard Bank of SA Ltd} 1999 4 SA 928 (SCA) at para 28.
\item \textsuperscript{216} \textit{Mort} (note 215 above) at 474–475.
\item \textsuperscript{217} Ibid at 475.
\end{thebibliography}
Soon thereafter, the court in Coetzee also endorsed the communitarian understanding of human dignity as a moral freedom which entails a duty to respect the human dignity of others when it held that a contract term that infringed upon one of the contracting party’s human dignity would be considered contrary to public policy. Woolman emphasised that the court’s judgment was based on Kant’s second categorical imperative as the court stated that the contract infringed upon the soccer player’s human dignity because in terms of the contract he was ‘helpless’ and ‘treated just like an object’, and hence, not treated as an end in himself. Specifically, the court held that the National Soccer League (NSL) rules as incorporated into the contract treated the ‘players as goods and chattels who are at the mercy of their employer once their contract has expired’. Therefore, the court’s definition of what it means to treat a person as an end in herself accords with Wood’s interpretation of Kantian dignity that endorses the idea that treating a person as an end in herself denotes respect for her unique set of ends which would include that she should not be deprived of control over her life; an understanding which was also endorsed outside the field of contract law by the Court in Pillay.

However, in Brisley, dealing with the enforceability of a non-variation clause, the court stressed the fact that ‘public policy generally favours the utmost freedom of contract’ and struck a damaging blow to the principle of good faith when it held that good faith was an abstract value underlying the substantive common law of contract and not an independent substantive rule that can be used to strike down a contract that would otherwise be enforceable. If this would be allowed, the court argued, it would mean that whether a contract would be enforceable or not would depend on what a particular judge viewed as fair and just. The court held that granting a judge such a discretion would ignore the principle of pacta sunt servanda and lead to commercial uncertainty. In this respect, the court expressly rejected Olivier J’s minority judgment in Saayman. Finally, it held that the principles in Sasfin cannot be used to prevent the enforcement of contractual terms that are not in themselves contrary to public policy. Further, that even if the principles in Sasfin could be extended to the unfair enforcement of terms, it should only be applied in exceptional cases.

Although the court did not recognise good faith as an open norm, it did not hesitate to rely on public policy to deal with contractual unfairness, the latter also being an open-ended

---

218 Coetzee v Comitis & Others 2001 1 SA 1254 (C)(‘Coetzee’) at paras 34 and 41. See also Lubbe (note 19 above) at 422.
219 Coetzee (ibid) at para 34.
220 Woolman (note 80 above) at para 36.4(f). See also Lubbe (note 19 above) at 422.
221 Coetzee (note 218 above) at para 38.
222 Compare the discussion in the text at note 78 above.
223 Compare the discussion of Pillay (note 121 above) in the text at note 167 above.
225 Brisley (note 224 above) at paras 21 and 22 (Court rejected the views in Mort (note 215 above).)
226 Brisley (note 224 above) at para 24.
227 Ibid.
228 Ibid at paras 14–22.
229 Ibid.
230 Lubbe (note 19 above) at 397. See also Bhana & Pieterse (note 1 above) at 873.
concept. In addition, as the court followed *Sasfin*, its decision can be seen as an endorsement of the communitarian understanding of Kantian dignity that supports a moral freedom (human dignity as empowerment) and the resultant duty to respect the human dignity of others (human dignity as constraint) which results in a balancing act between the interests of the parties as reflected in the contract terms themselves. As in *Sasfin*, the court also limited the investigation of fairness to the contractual terms themselves; and did not take account of the socio-economic circumstances of the parties at conclusion or enforcement of the contract. This resulted in the court emphasising that ‘public policy favours the utmost freedom of contract’ which reflects the Kantian external realm of right that requires the maximisation of everyone’s freedom and does not promote socio-economic rights or substantive equality. In light of the Constitution’s commitment to social transformation, the court’s decision was met with criticism.

Soon thereafter, the enforceability of an exemption clause that excluded a private hospital’s liability for the negligent conduct of its nursing staff was considered in *Afrox*. With reference to *Sasfin* the court accepted that a contractual term which is so unfair as to be against public policy would be unenforceable, but cautioned that the power to declare contracts contrary to public policy should be exercised sparingly and only in the clearest of cases to prevent the decision being based on the judge’s own subjective perception of fairness. This can be viewed as an endorsement of the communitarian understanding of Kantian dignity that reflects a moral freedom and the resultant duty to respect the human dignity of others that results in a value judgment (expressed through the concept of public policy) and requires the court to balance the interests of the parties to determine the fairness of the contract.

The court held that the values informing public policy are rooted in the Constitution and its founding values, namely dignity, equality and freedom. By quoting from appeal judge Cameron’s concurring judgment in *Brisley*, the court stated that these values require the courts to show restraint when striking down contracts because contractual autonomy is part of the constitutional value of freedom, and ‘[s]horns of its obscene excesses’, contractual autonomy informs the constitutional value of dignity. So similar to its approach in *Brisley*, the court emphasised the empowerment conception of human dignity in a way which results in the maximisation of the individual’s freedom as reflected in the communitarian understanding of Kantian dignity. In doing so, the court further elevated contractual freedom to the status of a constitutional value. However, the court did admit that the unequal bargaining position of the parties is a relevant factor in deciding whether a contract term is contrary to public policy, but argued that its presence alone will not result in the contract term being contrary to public policy. The court thus recognised that the promotion of socio-economic rights and substantive equality could be taken into account when determining the fairness of a contract.

---

232 Ibid at para 18.
234 Ibid.
235 Ibid at para 18.
236 Ibid at para 22 quoting *Brisley* (note 224 above) at para 94.
237 Ibid at para 23. See also Bhana & Pieterse (note 1 above) at 877–878; Lubbe (note 19 above) at 415.
238 Afrox (note 233) at para 12.
term, albeit in a very limited way. Again, the court’s decision was met with criticism. Finally, the court confirmed that good faith is an abstract value underlying the substantive common law of contract and not an open term that can be used to strike down a contract that would otherwise be enforceable.

V MAKING SENSE OF BARKHUIZEN, BREDENKAMP AND BOTHA

A Barkhuizen

The issue of contractual fairness finally arrived in the Court in Barkhuizen which dealt with the enforcement of a time-limitation clause in a short term insurance contract. The Court held that constitutional challenges to contractual terms must be determined by testing the terms against public policy, which in turn is informed by the Constitution and the constitutional values of freedom, dignity, equality and the rule of law. A contractual term that is contrary to such values is contrary to public policy and unenforceable.

The Court referred to the test laid down in Sasfin which entails a balancing act between the policy considerations of freedom and sanctity of contract on the one hand and simple justice between individuals on the other. It held that such a balancing act necessarily implicates notions of fairness, justice, equity and reasonableness. The Court held that a balance must be found between ‘unacceptable excesses’ of contractual autonomy, and allowing ‘individuals the dignity and autonomy of regulating their own lives.’ The Court further expressly rejected the appeal court’s contention that the fact that a contractual term is unfair or might operate harshly cannot lead to the conclusion that the term is contrary to constitutional values and principles. These statements can be viewed as a confirmation of the communitarian understanding of Kantian dignity as a moral freedom and the resultant duty to respect the human dignity of others. It further reflects an acceptance that this approach to human dignity necessarily results in a balancing act between the interests of parties as reflected in the contractual terms themselves through a value judgment based on fairness.

However, the Court then laid down a two-part test for determining fairness which resulted in some uncertainty as to the above proposition. The first part of the test concerns the fairness of the clause itself and requires a balancing act between the policy considerations of freedom and sanctity of contract which gives effect to the constitutional values of freedom and human dignity on the one hand, and another policy consideration as reflected in a constitutional right or value (in casu the right to access of justice) in support of the non-enforcement of the

---

240 Afrox (note 233 above) at para 32.
241 Ibid at paras 28–29.
242 Ibid at para 50, fn 33 referring to Sasfin (note 189 above) at 9.
243 Barkhuizen (note 3 above) at paras 51 and 73.
245 Barkhuizen (note 3 above) at para 72.
246 Cf the discussion of Sasfin in the text at note 195 above.
247 Barkhuizen (note 3 above) at para 56.
contract on the other.  

This examination is objective in nature as it deals with these values on an abstract level as reflected in the terms of the contract itself.  

To some extent this approach can still be interpreted as aligning with the Sasfin test based on a communitarian understanding of Kantian dignity as explained above. However, it is unfortunate that the Court endorsed the idea of human dignity as empowerment only when it reiterated that contracts freely and voluntarily entered into must generally be enforced. It did not refer to or consider the idea of human dignity as constraint as reflected in the duty to respect the human dignity of the other contracting party which could be used to argue for the non-enforcement of the contract. In other words, the Court did not recognise that the constitutional value of, or right to, human dignity can also be invoked on the other side of the balancing scale to limit contractual freedom and that this would necessarily implicate the notion of fairness because it entails a balancing act of the different parties’ interests as reflected in the contractual terms themselves. The Court’s failure to consider the idea of human dignity as constraint created legal uncertainty as to whether a court had an equitable discretion to set aside unfair contract terms or if a party arguing for the non-enforcement of the contract always has to implicate a specific fundamental right before fairness becomes relevant.  

The Court further held that if the clause objectively does not violate public policy, it must be determined whether the clause itself violates public policy in light of the relative situations of the contracting parties, which would include an assessment of the bargaining positions of the parties. This determination is subjective in nature. Therefore, the Court’s reasoning reflects a conceptualisation of human dignity as constraint which denotes not only a duty to respect the human dignity of others but also to promote socio-economic rights and substantive equality in private dealings with others. This is an extension of Sasfin in which the determination of fairness was limited to the contract terms themselves. As Sasfin is based on an understanding of Kantian dignity which struggles to justify and promote socio-economic rights and substantive equality, it can be argued that this extension was inspired by ubuntu. As was shown above, human dignity through ubuntu results in a duty on the individual to promote the socio-economic wellbeing of the other community members. This is further supported by the fact that the Court recognised that public policy as an open norm should  

---


249 Barkhuizen (note 3 above) at para 59 (Court refers to the ‘objective terms’ of the contract). See also Hawthorne (note 248 above) at 398; Botha (note 36 above) at 212.  

250 Compare the discussion in the text at note 246 above.  

251 Brand (note 19 above) at 85; Bhana (note 19 above) at 274; Lubbe (note 19 above) at 420–421.  


253 Barkhuizen (note 3 above) at para 59.  


255 D Bhana & N Broeders ‘Agreements to Agree’ (2014) 77 Tydskrif vir die Hedendaagse Romeins-Hollandse Reg 164, 175; Cornell & Muvangua (note 116 above) at 24; Hawthorne (note 248 above) at 400.  

256 Compare the discussion on the differences between Kantian dignity and ubuntu in the text at note 154 above.
be ‘informed by the concept of ubuntu’.

Thus, the Court expanded the limited scope of substantive equality considerations as set out in Afrox and this development is the result of a harmonisation between Kantian dignity and ubuntu.

The second part of the test for fairness investigates whether, in spite of the fact that the clause itself does not violate public policy, enforcement of the clause would be fair in light of the circumstances which prevented compliance with it. Again, the second part of the test is subjective in nature and promotes substantive equality. Accordingly, it is also inspired by ubuntu. It is further an extension of the fairness test in Sasfin as it is concerned with the unfair enforcement of a contract term, and not only whether the contract terms themselves are fair or not. Again, this reflects a conceptualisation of human dignity as constraint which denotes not only a duty to respect the human dignity of others but also to share in the responsibility of creating the necessary conditions for the realisation of others’ human dignity through the promotion of substantive equality. This development can also be viewed as a harmonisation between Kantian dignity and ubuntu.

Finally, the Court obiter questioned the restricted role of good faith in the common law of contract.

B Bredenkamp

The Supreme Court of Appeal in Bredenkamp attempted to clarify the legal uncertainty created in Barkhuizen. This case dealt with the exercise of a contractual right that entitled the bank to close a client’s bank accounts on reasonable notice and for any reason. Relying on Barkhuizen, the client argued that the bank was required to exercise this right fairly and for good cause. In other words, without implicating a constitutional value or right in support of the non-enforcement of the contract, the client relied on the second part of the test for fairness in Barkhuizen that deals with the unfair enforcement of a contract term.

The court stated that the judgment in Barkhuizen did not hold that the enforcement of a valid contractual clause must be fair and reasonable where no public policy consideration in the Constitution or elsewhere is implicated. Therefore, fairness and reasonableness become relevant only when a specific constitutional value or right is implicated and it must be determined whether the clause or its enforcement is contrary to public policy. As indicated by the court in Barkhuizen v Napier, this balancing act will be between freedom and sanctity of contract that gives effect to the constitutional values of freedom and human dignity, on the one hand, and another policy consideration as reflected in a constitutional right or value

---

257 Barkhuizen (note 3 above) at para 51.
258 Compare the discussion of Afrox in the text at note 238 above.
259 Barkhuizen (note 3 above) at para 56.
260 Ibid at para 59. See also Wallis (note 254 above) at 552–553; Bhana & Meerkotter (note 11 above) at 504; Bhana (note 252 above) at 509.
261 See again the sources listed in note 255 above.
262 Barkhuizen (note 3 above) at para 82.
263 Bredenkamp (note 4 above) at para 6.
265 Compare the discussion in the text at note 259 above.
266 Bredenkamp (note 4 above) at para 50.
267 Ibid at paras 43–44.
268 Barkhuizen (note 3 above) at para 57.
in support of the non-enforcement of the contract, on the other. As mentioned before, this balancing act is objective in nature.\textsuperscript{269} This means that a court must first identify the relevant policy considerations in support of the non-enforcement of the contract as reflected in specific constitutional values and/or rights before it may apply the second part of the test for fairness which is subjective in nature.\textsuperscript{270}

By following the Court’s approach in Barkhuizen, the appeal court recognised that human dignity can be invoked in support of contractual autonomy and sanctity as reflected in its empowerment conception to argue for the enforcement of the contract. However, it also failed to recognise that human dignity as a constitutional value or right can also be invoked on the other side of the balancing scale in support of the non-enforcement of the contract and that this would necessarily implicate the notion of fairness. Due to this failure, the appeal court rejected the idea that fairness is an overarching requirement in the common law of contract,\textsuperscript{271} and expressly stated that an equitable discretion cannot be allowed because it would defeat the rule of law entrenched as a founding constitutional value.\textsuperscript{272}

This approach allowed the Court to avoid applying the subjective test and hence it did not take account of the socio-economic position of the parties and the issue of substantive equality between the parties was not addressed. Consequently, the Bredenkamp decision was criticised because it endorsed a very liberal conception of human dignity and freedom and did not promote social justice.\textsuperscript{273} Although the Court accepted that contractual freedom is not an unconstrained freedom, the court emphasised the empowerment conception of human dignity which results in the maximisation of the individuals’ freedom to contract.\textsuperscript{274} Thus, the court’s approach aligns more with the communitarian understanding of Kantian dignity than an understanding of human dignity informed by ubuntu that requires a greater limitation on an individual’s freedom in order to promote substantive equality in private dealings. This is supported by the fact that the Bredenkamp court made no reference to ubuntu.

It could be argued that the court’s failure to consider how human dignity is understood through ubuntu, resulted in Yacoob J’s call for ubuntu to inform the notion of public policy in the common law of contract in Everfresh.\textsuperscript{275}

C. Botha

As mentioned in the introduction, the Court’s decision in Botha was criticised because it created legal uncertainty as to the role of fairness in the law of contract.\textsuperscript{276} In this article, I try to resolve some of this legal uncertainty by arguing that the Court’s decision reflects an

\textsuperscript{269} Compare the discussion in the text at note 249 above.

\textsuperscript{270} The second part of the test states that if the clause itself is reasonable and does not violate public policy, it should be determined whether the clause should be enforced in light of the circumstances which prevented compliance with the clause.

\textsuperscript{271} Bredenkamp (note 4 above) at paras 50–53.

\textsuperscript{272} Ibid at para 39 referring to s 1(c) of the Constitution. The role of the constitutional value of the rule of law as relating to contractual fairness falls outside the scope of this article.

\textsuperscript{273} Bhana (note 252 above) at 515.

\textsuperscript{274} Ibid.

\textsuperscript{275} Everfresh (note 8 above) at para 23.

\textsuperscript{276} Compare the discussion in the text at note 9 above. See also L Hawthorne ‘Rethinking the Philosophical Substructure of Modern South African Contract Law: Self-actualisation and Human Dignity’ (2016) 79 Tydskrif vir die Hedendaagse Romeins-Hollandse Reg 286; Wallis (note 254 above) at 554–557.
understanding of human dignity as both empowerment and constraint, the first denoting a moral freedom and the latter reflected in the duties on an individual to respect the human dignity of others and to promote the socio-economic conditions for the realisation thereof.

The facts in Botha concerned the enforcement of a cancellation clause in an instalment sale of immovable property between Botha (as buyer) and a trust (as seller). The cancellation clause provided for the cancellation of the agreement and the forfeiture of all sums already paid by Botha in the event of a breach by her. The contract also provided that Botha could demand transfer of the property in terms of s 27 of the Alienation of Land Act 68 of 1981 (ALA) after she had paid at least half the purchase price. After Botha had paid three quarters of the purchase price, she began to default on the payments and the trust sued for cancellation and eviction. In turn, Botha demanded transfer of the property in terms of s 27 of the ALA as provided for in the contract. One of Botha’s main contentions was ‘that the enforcement of the cancellation clause, where more than half the purchase price (had) been paid, and in the face of a demand for a transfer pursuant to s 27, (was) contrary to public policy’. In determining this question, the Court held that public policy generally requires enforcement of contractual obligations freely and voluntarily undertaken and that this gives effect to the constitutional values of freedom and human dignity. This is an endorsement of human dignity as empowerment that supports the enforcement of the contract.

In support of her case, Botha contended that the enforcement of the cancellation clause by the trustees was contrary to public policy because it violated her constitutional rights. It is not stated which constitutional rights Botha implicated in her application before the Court but most likely she relied on her rights to dignity and equality as these were the rights cited before the Supreme Court of Appeal. Therefore, it can be argued that Botha invoked human dignity as the specific constitutional right or value in favour of the non-enforcement of the contract term as required by the Court in Bredenkamp. Furthermore, by linking human dignity and equality, the idea of human dignity as constraint as endorsed by the Court refers not only to a duty on the individual to respect the human dignity of another but also the duty to promote the realisation of the latter’s human dignity through the promotion of substantive equality.

The Court then applied this idea of human dignity as constraint which it conceptualised through the principle of good faith:

The principle of reciprocity falls squarely within this understanding of good faith and freedom of contract, based on one’s own dignity and freedom as well as respect for the dignity and freedom of others. Bilateral contracts are almost invariably cooperative ventures where two parties have reached a deal involving performances by each in order to benefit both. Honouring that contract cannot therefore be a matter of each side pursuing his or her own self-interest.

277 Botha (note 9 above) at para 4.
278 Ibid at paras 5–6.
279 Ibid at para 8.
280 Ibid at para 19.
281 Ibid at para 23.
282 Ibid at para 19.
283 Ibid at para 15.
284 Bhana & Meerkotter (note 11 above) at 505 (Argue that the Court did not explicitly implicate an enumerated right as required by Bredenkamp.)
285 The idea of substantive equality is linked to human dignity and equality. Botha (note 9 above) at para 28.
without regard to the other party’s interest. Good faith is the lens through which we come to understand contracts in that way.\textsuperscript{286}

Although this statement was made in respect of the application of the \textit{exceptio non adimpleti contractus} to the facts of the case, the Court later stated that the enforcement of the cancellation clause and the concomitant forfeiture of the purchase price already paid was unfair\textsuperscript{287} for the same reasons because it constituted a disproportionate penalty in circumstances where three-quarters of the purchase price had already been paid.\textsuperscript{288}

The Court’s decision can therefore be viewed as an application of the test for fairness as set out in \textit{Barkhuizen} and as further expanded upon in \textit{Bredenkamp}. By referring to both parties’ human dignity\textsuperscript{289} the Court recognised that determining the fairness of a contractual term or the enforcement thereof will always entail a balancing act between the human dignity of the contracting party who wants to enforce the contractual term as expressed through the maxims of freedom and sanctity of contract and the human dignity of the other contracting party who avers that the contractual term or its enforcement infringes her human dignity as expressed in the demand that a human being should never be treated merely as a means to an end, but always also as an end. This means that the unique set of ends of both parties must be balanced which entails that the court has to weigh up the interests of both parties in order to determine whether the contractual term or its enforcement constitutes an unfair infringement of one of the party’s interests and hence violates her human dignity. Consequently, by impliedly invoking the concept of human dignity as constraint through the principle of good faith the Court was able to apply the subjective test for fairness in determining whether the enforcement of the cancellation clause was unfair.

In coming to the decision that the enforcement of the cancellation clause was unfair, the Court took specific notice of the fact that the cancellation in the specific circumstances of the case would entail the forfeiture of almost three-quarters of the purchase price that had already been paid by Botha while the trust would keep the property. The Court thus balanced the interests of the parties and came to the conclusion that the enforcement of the contract term would be unfair.\textsuperscript{290}

As the Court linked the concept of good faith to both duties as reflected in human dignity as constraint, namely to respect and promote the realisation of the human dignity of the other contracting party through substantive equality, the decision relies by implication on ubuntu. As was argued above,\textsuperscript{291} the subjective part of the test for fairness which includes considerations based on substantive equality is an expression of ubuntu because it requires an investigation into the contracting parties’ actual social and economic conditions. Accordingly, ubuntu can

\textsuperscript{286} \textit{Botha} (note 9 above) at para 46.
\textsuperscript{287} The court specifically speaks of the ‘fairness of awarding cancellation’. \textit{Botha} (note 9 above) at para 51.
\textsuperscript{288} Ibid.
\textsuperscript{289} Ibid at para 46 (Court speaks of the ‘reciprocal recognition of the dignity … of the respective contracting parties’.)
\textsuperscript{290} See also S Woolman ‘On the Reciprocal Relationship between the Rule of Law and Civil Society (2015) \textit{Acta Juridica} 374 (Contends that a necessary precondition for any viable system of law requires that each citizen recognize all other citizens as worthy of ‘equal concern and respect’.)
\textsuperscript{291} I am not arguing that the public policy consideration of legal certainty is irrelevant in determining whether contract terms or their enforcement are contrary to public policy. However, such an investigation falls outside the scope of this article. For academic criticism that \textit{Botha} aggravated the existing uncertainty regarding the law, see Wallis (note 254 above) at 554–557.
\textsuperscript{291} Compare the discussions in the text at notes 255 and 261 above.
be construed as the subtext. Thus, it is unfortunate that the Court did not refer to ubuntu expressly as doing so could have enabled the Court to consider further factors in favour of the non-enforcement of the cancellation clause. For example, Hawthorne argued that the parties were in an unequal bargaining relationship and that Botha had ‘little negotiating power’ in concluding the contract.\(^{292}\) A further possible factor is the fact that Botha used the property to operate a laundry service through a closed corporation of which she was the sole member.\(^{293}\) Losing the property could well have resulted in her losing the business that provided her with a substantial part (if not the sole source) of her income.\(^{294}\) Therefore, enforcement of the cancellation clause, coupled with the forfeiture of the purchase price already paid, might well have resulted in the deprivation of Botha’s livelihood.

V CONCLUSION

The above analysis illustrates that the Court is in the process of developing both aspects of human dignity as empowerment and constraint in the common law of contract which is the result of the harmonisation of Kantian dignity and ubuntu. Human dignity as empowerment denotes a moral freedom while human dignity as constraint refers to the duty to respect the human dignity of the other party as well as the duty to promote the realisation of the other party’s human dignity through the idea of substantive equality. The contracting party wishing to enforce the contract term relies on freedom and sanctity of contract, while the other contracting party who wants to prevent its enforcement avers that the contractual term infringes her human dignity, as expressed in the demand that a human being should never be treated merely as a means to an end, but always also as an end. This approach inevitably requires a balancing act between the interests of the parties as reflected in the contractual terms themselves as well as taking account of the circumstances surrounding the conclusion and enforcement of the contract. Hence the notion of fairness is implicated when determining whether a contractual term or its enforcement is contrary to public policy.

Finally, it is therefore submitted that the proper appreciation of the constitutional value of human dignity in the common law of contract results in a general equitable discretion to declare contractual terms or their enforcement unfair and therefore invalid. Furthermore, it is possible to interpret the Court’s decision in Botha as creating such an equitable discretion by expressing the idea of human dignity as constraint through good faith.

---

292 Hawthorne (note 276 above) at 299.

293 Botha (note 9 above) at para 3.

294 The relevance of these factors are supported by Bhana & Meerkotter (note 11 above) at 505 (Argues that the facts of the case could have implicated the rights to property (s 25 of the Constitution) and freedom of trade, occupation and profession (s 22 of the Constitution).)
The More Things Change … ¹
Multiracialism in Contemporary South Africa

MARTIN BRASSEY

ABSTRACT: The system of multiracialism, by which the races in South Africa were streamed and administered separately, still prevails and, though proscribed by an avowedly non-racial Constitution, is applauded by our highest court. Nothing prevents us from embracing the constitutionally-mandated goal, and remedial action should invoke race only when necessity dictates.

KEYWORDS: race, multiracialism, non-racialism, apartheid, separate development, equality, ICERD, section 9, affirmative action, remedial action, quota

AUTHOR: BA (UCT) LLB (Wits) H Dip Tax Law; Visiting Professor of Law, University of the Witwatersrand; Senior Counsel. Email: brassey@counsel.co.za

ACKNOWLEDGEMENTS: I need to disclose an interest and so a potential bias: I acted for Solidarity in both the cases to which it is a party, though not in each of the courts in question. I also appeared in the Insolvency Practitioners case.

¹ ‘Plus ça change, plus c’est la même chose.’ (The more things change, the more they remain the same.)
I INTRODUCTION

South Africa is by law pronounced and proclaimed to be a non-racial state. The Constitution, the basic law of the country, solemnly states as much: in s 1, which is termed the founding clause, it expressly says that non-racialism is one of the central values of the single sovereign, democratic, state created by the instrument.

South Africans, taken generally, applaud this fact. For fifty years, they suffered under a system in which people were divided and ruled by race, and they now want no part of it. They renounce, indeed roundly denounce, the naked racism of apartheid and the sophisticated variant it spawned, separate development. No longer, they say, will we be divided up and governed on the basis of race. We have ceased to be black, coloured, Indian or white and have become individuals who are entitled to be treated as such. By our Constitution, we have committed ourselves to a non-racial future; this is what we believe in, what we have fought for, and what we have achieved. In the spirit of Ladysmith Black Mambazo, our country’s glorious *acapella* group, we believe we can now raise our voices and sing out: ‘Different colours mean nothing to me; different languages mean nothing to me; different names mean nothing to me.’

Or can we? In the name of equality, under the rubric of affirmative action, the state continues to treat people by reference to race and, using the self-same system of racial categorization by which apartheid operated, deals with them accordingly. The policy, which is widely implemented, is nowhere more rigorously applied than in the field of employment, and the public service, especially but by no means exclusively. In the interests of ‘demographic representivity’, job applicants are uniformly characterised by colour and succeed, whatever their intrinsic merits, only if the vacancy is for a person taken to fall within the group in question. In form, this is not non-racialism; nor is it substantively so, since whites, who can scarcely be said to be the victims of past discrimination, benefit and suffer equally from the system. What it is, is multiracialism.

To boot, it is multiracialism of a special sort. It is multiracialism that shares precisely the same inspiration, the same philosophical justification and the same structure as the hated policy of separate development. We cannot, it seems, throw off our past: what plagued us once continues to haunt us still. Why this should be so raises a number of questions.

- Is there, we ask ourselves, something in the very notion of equality that forces us to employ multiracialism as our organizing model?
- Is there something in our legal tradition, our common law especially, that has this result?
- Are there imperatives of international law recognised by us as mandatory that demand compliance with the system?
- Finally, are there provisions in the Constitution – the Equality Clause not least – that override the aspirations in the founding clause so as to make multiracialism obligatory?

In this article I consider these questions and conclude that nothing forces us back into the clutches of multiracialism. Quite the contrary: the Constitution, in terms and in spirit, obliges us to embrace non-racialism and, by making race cardinal to our decision-making, we flout its basic tenets. In itself this is a bad thing, since the instrument, negotiated so earnestly between white and black, deserves our respect, even fealty. No less bad are the consequences, which I consider, somewhat fleetingly, before I finish the essay. There are a few countries in which

---

2 Joseph Shabalala, producer, ‘Kangivumango’, a track to be found on the record album ‘*The Very Best of Ladysmith Black Mambazo – Rain, Rain, Beautiful Rain*’ (2004), compilation by Ladysmith Black Mambazo.
multiracialism is implemented and, in one or two of them, the system has arguably produced good results, but South Africa is most certainly not among them. The policy of dividing people into races and ‘developing’ them separately has failed us in the past and it will not serve us in the future.

II  IS MULTIRACIALISM, SOMEHOW, AN INHERENT FEATURE OF EQUALITY?

Most, if not all, judgments entail a comparison of one sort or another. Sometimes it will be between things. If we say that something is hard, we have in mind its corollary, a thing that is soft. Sometimes it will be between people: when we say that X is the eldest of the children, we bring to mind the siblings she has and recognise that they are younger than her. Sometimes we mix up the two, people and things. If we say that someone lives in poverty, we set a standard by reference, first, to the material things required to support life and, secondly, to the way other people – richer people – are living. The fact that the first is absolute and the second is relative makes no difference for present purposes. The standard, once set, still acts as a basis by which we make the comparison.

The comparison, it should be obvious, entails a choice. Once we have the subject of our decision-making, we select the person or thing with which we will make the comparison and the criterion by reference to which the comparison will be made. The comparator need not be real, and often is not. A notional thing or person can be postulated as the basis of comparison – lawyers, accustomed to use such fictions as the reasonable person, know this well enough – and sometimes the process of abstraction can be so extended that we all but forget that a comparison is being made. Such is the position when a standard, set by reference to how people might properly be expected to behave, becomes the comparator idealised as a principle. Bound up in the rule that murder is the intentional and unlawful killing of another human being, for example, is the idea that, when two killers are compared, one may be condemned when the other is not.

In the absence of constraint, moral, legal or social, the choice is entirely ours to make. We are continually reminded that it is wrong to compare apples with pears, but this is so only if our concern is with the essential ‘appleness’ of the thing. If our concern is to know whether each share the quality of being a fruit, then it is entirely proper to make the comparison, in the process identifying the subject of the analysis (the apple), the comparator (the pear) and the criterion by which to compare them (the quality of being a fruit). In the same way, we can compare one person, older and taller though she is, by reference to, say, the colour of their hair or the way they laugh.

In the domain of equality, the self-same comparison is, and must be, made. The subject of the evaluation must be chosen, so must the comparator, and so too the criterion by reference to which they will be evaluated. The only difference is that the assessment is narrower; instead of seeking points of identity or disparity, which may be large or small, we are now asking whether the subject and comparator are, by reference to the applicable criterion, the same or not. The sameness can, in principle, be of any quality: if we choose weight as the criterion, for instance, we can say that a lump of lead is equal to a sack of feathers if each weighs a kilogram; if we are concerned with their texture, however, we must conclude that they are unequal. In the same way, a child may be unequal to her siblings in height but be their equal in appetite or generosity.
The opening scenes of King Lear provide a striking illustration of how the methodology of equality works. The eponymous king, now in his dotage, decides to unburden himself of his estates. He has three daughters and they, he proposes, should each be the recipients of his largesse. A courtier expects him to prefer the daughter whose husband he likes better, but the King takes up the stance, initially anyway, that each should benefit equally. Before making the final decision, however, he requires them to describe the depth of each of their love for him. The elder two play along, but the youngest will have none of it. Speaking from her heart, she says simply ‘I love your majesty according to my bond, no more nor less.’ In time, she tells him, she will be wed, and since her husband will ‘carry half my love with him,’ she cannot copy her sisters by pretending she will always love him alone. The King, enraged beyond measure, disowns her and grants the whole of his lands to the other two in equal shares. What ensues is terrible and, in the finest tradition of Shakespearean tragedy, none of the principal figures emerges alive.

In making the distribution, Lear proceeded – could not but proceed – in precisely the way we have foreshadowed. First he decided on the benefit to be conferred, then he selected the people whom he would benefit, and finally he resolved on the extent of the benefit. The people comprise his daughters as distinct from their husbands, his courtiers or the public at large: the benefit consists of his lands in their totality, not just a part of them; and the distributions are to be made equally provided the threshold requirement – the proper expression of filial piety – is satisfied. The basic comparison determines the set of potential beneficiaries by selecting in favour of the children and against everyone else; but within the set is a subset, determined by reference to the expression of love, that ultimately serves to favour the eldest two daughters over the youngest.

In the tale, we discover not just how decision-making works – the process or, if you like, the methodology. We also see that, unless some constraint dictates otherwise, the process is wholly open-ended. Who is to be compared to whom is a matter of discretion; so too is the criterion by reference to which the comparison will be made and the outcome of the comparison (will there be an equal distribution or one that is proportioned?) is a matter of discretion. How the discretion is exercised is always a matter of who we choose to compare with whom, and what we elect to use as the criterion for comparison. The courtier expected that the husbands would be the subject of comparison, that the king’s relative love and affection would be the criterion, and that the lands, which were the object of the decision, would be distributed proportionately. In fact, the daughters were compared on the basis of their expression of love for the kings and the distribution was made equally between two of them.

Until populated in this way, equality is and must always be an abstraction, a purely formal concept that can tell us nothing of substance. Peter Westen is right, therefore, when he describes the concept as an ‘empty idea’:

\[5\] It deserves this description because it is ‘empty of content’, ‘an empty form having no substantive content of its own.’ There is no such thing as ‘substantive’ equality.

---

3 King Lear Act I Sc 1.
4 Ibid.
5 As so often, this is a metaphor that conceals as much as it reveals. Beneath the surface are two ideas: first, that equality is meaningless until parameters are generated, invariably by recourse to normative standards, to give it content; secondly, that it remains empty even once the parameters have been selected, since recourse to the underlying values axiomatically makes the canon of equality otiose. What I am concerned with is the first, not the second conception; my interest is in the indeterminacy of equality as a concept.
equality until we have decided who should be equal to whom, in what respects, and to what degree. There is, as a result, no point in peering into the concept in the hope that it will, in and of itself, yield eternal verities about the way society should be structured.7 ‘Without moral standards,’ Westen says, ‘equality remains meaningless, a formula which can have nothing to say about how we should act.’8 There is no closed list of the standards we can invoke and so no closed list of the parameters we can employ, so it is ‘a mistake to believe that the various kinds of equality can be enumerated as a multiple yet finite list.’9 In the words of Lakoff, ‘[o]n the one hand, there are as many substantive versions of equality as there are substantive notions of right and entitlement by which persons can be said to be “alike” or “unalike”; on the other hand, there is only one formal idea of equality – that “likes should be treated alike”’.10

There is, in principle, nothing that drives us to employ one set of standards rather than another. We make our choices by responding to our conceptions of right and wrong, good and bad, useful and useless, or whatever. Since we share such values to a greater or lesser degree, we are quick to take the parameters as given and invoke equality as though, in itself, it were instructive. We say people should be equally treated and when asked how, we say, according to merit, or needs, or works, or effort, or wants.11 But these, to repeat, are simply the choices that commend themselves to the moral conceptions of the proponent; absent a system of values, they have no inherent justification, and most certainly cannot be validated by proclaiming the merits of equality. Without deciding who should be equal to whom and in what respects we can resolve nothing. We cannot say whether sisters should be equal to their brothers, women should be equal to men, or blacks should be equal to whites.12

The doctrine of equality, it follows, makes no demands that we divide people by reference to the colour of their skin. If we can and do make this division, we make a choice that is ours. By parity of reasoning, the doctrine makes no demands that, once people are so divided, the resulting groups should, in some respect or other, be equally treated. If we can and do decide that they should, the choice once more is ours. Nothing intrinsic to the concept of equality can ever determine the question; its answer must ever be sought in values that are extrinsic to it and employed to make it function. No doubt some considerations appear more compelling than others: for example, our conceptions of the proper connection between cause and effect might suggest, at least to people of a liberal persuasion, that race is an extraneous factor in determining distributive social and economic questions. These are, however, just expressions of our ideas and values, which may or may not be dominant at any one time, and in no way derogate from the general proposition that is now being advanced.

7 This is no idle point. There are writers who propound just such a notion: for just one of many examples, see C Albertyn & Beth Goldblatt ‘Equality’ in S Woolman & others The Constitutional Law of South Africa 35-5.
8 Westen (note 6 above) 547.
9 Ibid 540 n8.
10 Ibid.
11 Jennifer L Hochschild What’s Fair (1981) usefully identifies the criteria typically invoked in support of distributive justice and places them on a continuum from ‘equality to differentiation’: See 51-51. The continuum proceeds from Strict Equality (eg per capita), through Need, Investments (how much in time effort or money has been invested) and Results to arrive, finally, at Ascription (ie family, status etc.). Her book takes a random sample of 28 working adults and solicits their views on how rewards and benefits should best be allocated. Her research reveals: ‘Support for equality is weak in discussions of property rights, but grows stronger as the resource to be distributed shifts to social policies, pure political rights and authority’ (At 190).
12 The notion that equality, until populated, is a purely formal concept is one of the central themes of Amartya Sen’s prodigious work: see, in particular, The Idea of Justice (2009) ch 14.
Once this is understood, we are obliged to accept that multiracialism is just one more choice. As a matter of social policy, its merits can be debated, but nothing rooted in the principle of equality compels us to accept it as axiomatic. If we plump for it, either in itself or in one or other of its manifestations such as separate development or demographic representation, we do so because we believe this is justified by deeper values we hold dear.

III  DO OUR LEGAL TRADITIONS PROPEL US TOWARDS MULTIRACIALISM?

Under the common law, a person is free to act, or decline to act, unless the law prescribes otherwise. Freedom is the basic postulate of our law, and it is enjoyed by everyone to the extent that it is not abridged by law. Primary legislation enshrining race-based discrimination is not lightly presumed to create such an abridgment since it ‘is a wholesome rule of our law which requires a strict construction to be placed upon statutory provisions which interfere with elementary rights.’\(^\text{13}\) In dealing with subordinate legislation, the courts were wont to exercise greater powers of control. They regularly struck down regulations, ordinances and similar enactments if they deemed them to be unreasonable. Ordinarily, a rule would be unreasonable if it fell outside the range of rules that reasonable people might make, but within this context, the word was assigned a special, more limited, meaning. It was to denote something less than conventional unreasonableness but more than mere irrationality. Measures, by reference to this intermediate standard, would be struck down if they were ‘partial and unequal in their operation as between different classes; if they were manifestly unjust; if they disclosed bad faith; if they involved such oppressive or gratuitous interference with the rights of those subject to them as could find no justification in the minds of reasonable men.’\(^\text{14}\) If the measure was unjust but not manifestly so, it would survive scrutiny; so too if it disclosed some degree of justification even though the justification was not compelling. Parliament, the representative of the people, could sanction measures that were misguided or mistaken,\(^\text{15}\) but it could not be taken to legitimate the perverse or absurd. If a measure was ‘off the wall’, as the saying goes, then ‘the Court might well say, “Parliament never intended to give authority to make such rules; they are unreasonable and ultra vires.”’\(^\text{16}\)

Using the same interpretative approach, the common law courts would presume, unless the contrary was clear, that the lawgiver intended to promote the public interest by means properly tailored for the purpose. Unless constrained to do otherwise, they would reject a construction which produces a result that is, by prevailing standards, seen as arbitrary, partial or unequal in its operation. They would ‘not lightly construe a statute in such a way that its effect is to achieve apparently purposeless, illogical and unfair discrimination between persons who might fall within its ambit. [and, if] the language of the statute is reasonably capable of an interpretation which avoids that result, that is the interpretation which the court will give it rather than the one which would attribute to the legislature a whimsical predilection for purposeless and unfair

\(^\text{13}\) Dadoo Ltd & others v Krugersdorp Municipal Council 1920 AD 530 552.

\(^\text{14}\) Kruse v Johnson [1898] 2 QB 91 at 99–100. The dictum is treated as equally applicable in our law, as a long line of cases testifies.

\(^\text{15}\) A measure ‘is not unreasonable merely because particular judges may think it goes further than is prudent or necessary or convenient, or because it is not accompanied by a qualification, which some judges may think ought to be there’. Kruse (note 14 above) 100.

\(^\text{16}\) Ibid at 99–100. The applicability of Kruse in our law has never been questioned; Broome JP rightly remarked in R v Jopp 1949 (4) SA 11 (N) 13 that the doctrine has been approvingly applied in a ‘long line of cases’. 
discrimination.” Rationality was the touchstone and, in defence of liberty, they would, when they could, discountenance statutory constructions that were redolent of perverseness, injustice and, most pertinently for our purpose, unfairly discriminatory.  

Beyond a certain point, however, they were unable to proceed. Constrained by the doctrine of parliamentary sovereignty, they felt bound to implement statutes, however oppressive they might be, if they were unambiguous in the powers they conferred on officials or the burdens they imposed on subjects. This was a consequence of a conception of democracy that treated the will of Parliament, once properly discerned, as supreme and inviolable, recognizing the existence of no power of judicial review that might vest in the courts. But within those limits, they made themselves the protectors of the rights and liberties of the ordinary person as best they could. So, for instance, in Metal & Allied Workers Union v Minister of Manpower, the court held that, in the absence of an express power to make a racially-conditioned decision, the Registrar of Trade Unions acted unlawfully in registering non-racial unions as capable of representing black workers alone.

What the common law reprobated, it must be stressed, was not the use of race per se, but its use in an arbitrary or uneven way. In R v Carelse and R v Abdurahman, the court intervened not because the regulatory measures employed the criterion of race to segregate the amenities in question – beaches in the first case, railway carriages in the second – but because the actual division operated to the palpable disadvantage of the races who were not white. Discrimination itself was not the perceived mischief, but discrimination producing an unfair outcome. The consequence of this approach was an in-principle acceptance of racial streaming – a process we should term, and rightly term, multiracialism – and its condemnation only when its impact proved in fact to be disparate. The case on point, more infamous than celebrated, is Minister of Posts and Telegraphs v Rasool, in which our highest court was required to decide on the validity of a by-law that, without expressly being so authorised, segregated the facilities of the post office between ‘Europeans’ and ‘Non-Europeans’. In argument, the objecting party, an Indian individual, accepted that the two streams were functionally equivalent but claimed that the measure was inherently demeaning. The argument was rejected by a majority of the appeal court, whose views are captured in the judgment of De Villiers JA. ‘In my opinion … a discrimination which is not accompanied by inequality of rights, duties, privileges or treatment, is not per se unreasonable merely because it is made on grounds of race or colour’ Gardiner AJA dissented. Commencing with a useful discussion of the distinction between differentiation and discrimination, he went on to say: ‘I cannot shut my eyes to the fact that the instruction is actuated by the circumstances that a large number of Europeans object to being brought into

---

18 Of course, the common law judges were far from uniform in their application of these principles. It would be naïve to suggest that they might be. For an interesting insight into how they went about in the years between the world wars, see WHB Dean ‘Reason and Prejudice: The Courts and Licensing Bodies in the Transvaal’ in E Kahn Fiat Iustitia: Essays in Honour of Oliver Denys Schreiner (1983) 211. The author says: ‘it is quite clear that, although reluctant to recognise explicit racism, the courts were not prepared to tolerate decisions based purely on … race’. (At 214.)
19 1983 (3) SA 238 (N).
20 1943 CPD 242.
22 1934 AD 167.
23 Ibid at 182.
contact in public offices with non-Europeans, and that they regard the latter as being of a lower order of civilization’ so ‘any fresh classification on colour lines can … be interpreted only as a fresh instance of relegation of Asiatics and natives to a lower order … Such treatment is an impairment of the dignitas of the person affected’. This followed ineluctably from the fact that ‘the Asiatic is treated by our Legislation as being below the white man, [and] the native is treated as … being of yet a lower order.’

Would the outcome have been different had the judges been able to test the provision under a justiciable Bill of Rights? It is hard to say. The experience of the United States, which of course does have such a testing right, is equivocal. On the one hand, we have the late-nineteenth century decision of the US Supreme Court in *Plessy v Ferguson* endorsing the doctrine of multiracialism; on the other hand, we have *Brown v Board of Education of Topeka*, decided 50 years later, in which multiracialism was rejected and *Plessy* was explicitly overruled.

In *Plessy*, the court had to decide whether railway carriages could legitimately be segregated by race. The finding of the court was that a ‘statute which implies a legal distinction between the white and coloured races – a distinction which is founded on the colour of the two races, and which must always exist so long as white men are distinguished from the other race by colour – has no tendency to destroy the legal equality of the two races.’ Since it had been accepted, for the purposes of argument, that the two sets of carriages were equally commodious, the court declined to intervene. Separation, provided it was equal, was unexceptionable. Segregation was legitimate.

Justice Harlan, the sole dissenting voice, roundly rejected this thinking, holding that ‘State enactments regulating the enjoyment of civil rights upon the basis of race … under the pretense of recognizing equality of rights can have no other result than to render permanent peace impossible, and to keep alive a conflict of races, the continuance of which must do harm to all concerned.’ ‘The sure guaranty of the peace and security of each race,’ he continued, ‘is the clear, distinct, unconditional recognition … of every right that inheres in civil freedom, and of the equality before the law of all citizens … without regard to race.’ Even were this not so, he proceeded, the exercise of a power to segregate within the circumstances then prevailing would be bad in law. ‘Everyone knows that the statute in question had its origin in the purpose not so much to exclude white persons from railroad cars occupied by blacks, as to exclude coloured persons from coaches occupied by or assigned to white persons.’ The effect was to put ‘the brand of servitude and degradation upon a large class of our fellow citizens, – our equals before the law.’ ‘The thin disguise of “equal” accommodations for passengers in railroad coaches will not mislead anyone, nor alone for the wrong this day done.’

The argument that multiracialism, neutral in theory, is nevertheless discriminatory in practice proved determinative in *Brown*. The issue, said the court in the case, was this: ‘Does

---

24 Ibid at 190-191. The dialectic reflects underlying social attitudes and structures of power in a way that eerily tracks the seminal decision of the US Supreme Court in *Plessy v Ferguson* 163 US 537 (1896).
25 *Minister of Posts and Telegraphs v Rasool* (note 22 above) at 191.
26 163 US 537 (1896).
28 163 U.S. 537, 543.
29 Ibid at 561.
30 Ibid at 560.
31 Ibid at 557.
32 Ibid at 562.
the segregation of children in public schools solely on the basis of race, even though the physical facilities and other “tangible” factors may be equal, deprive the children of the minority group of equal educational facilities?\textsuperscript{33} In \textit{Plessy}, the majority had held it fallacious to assume that the enforced separation of the two races stamped the coloured race with the badge of inferiority and, should the contrary be thought true, this was “solely because the coloured race chooses to put that construction upon it.”\textsuperscript{34} In \textit{Brown}, the court held that this view represented a failure in “psychological knowledge” that had since been exposed by copious modern authority. The true position was that separate facilities are inherently unequal and, in consequence, the doctrine of “separate but equal” has no place within the public domain.\textsuperscript{35}

Cases such as these can be endlessly debated, but the debates are beyond this essay. For present purposes it is enough to observe that the common law, the fundamental source of our legal culture, is hostile to race-based programmes and tolerates them, if it tolerates them at all, only when the language of statutes makes a finding to the contrary untenable. We must conclude, then, that nothing in our legal traditions compels, or even encourages, our courts to embrace multiracialism. There are, to be sure, a few cases in which the courts endorsed race-based regulations in the absence of a statutory obligation to do so, but they are exceptional. \textit{Abdurahman},\textsuperscript{36} the seminal case on the common law, placed the issue of racial discrimination squarely within the rubric of reasonableness and held that such partial or unequal treatment is not to be countenanced unless specifically mandated by the enabling statute. Baxter is entitled to conclude, therefore that our common law courts, having adopted the \textit{dictum} in \textit{Kruse},\textsuperscript{37} evinced “an early abhorrence of discriminatory administrative action [by generally holding that] non-legislative discriminatory administrative action is unlawful if it is not clearly authorised by statute.”\textsuperscript{38}

\section*{IV DOES INTERNATIONAL LAW MAKE MULTIRACIALISM A REQUIREMENT?}

Section 39(1) of the Constitution stipulates that international law (which includes international treaties, conventions and customary norms) must be considered in the process of construing the Bill of Rights.

Of these sources, the most instructive is the International Convention on the Elimination of Race Discrimination (‘ICERD’), which entered into force in 1969, and a brief consideration of it will suffice for our purposes.\textsuperscript{39} In clause 4 it pertinently states that “[s]pecial measures taken for the sole purpose of securing adequate advancement of certain racial or ethnic groups or individuals requiring such protection as may be necessary in order to ensure such groups or individuals equal enjoyment or exercise of human rights and fundamental freedoms shall not be deemed racial discrimination.” The clause properly propounds a conception of substantive equality that is sensitive to past disadvantages and systemic patterns of discrimination.

\textsuperscript{33} Ibid at 493.
\textsuperscript{34} Ibid at 551.
\textsuperscript{35} Ibid at 494.
\textsuperscript{36} \textit{R v Abdurahman} 1950 (3) SA 136 (A) 149B–C.
\textsuperscript{37} Discussed above.
\textsuperscript{39} Their impact within this domain is well rehearsed in the UNESCO ‘Prevention of Discrimination: the Concept and Practice of Affirmative Action’ Final Report submitted by Mr Marc Bossuyt, special rapporteur in accordance with Sub-Commission Resolution 1998/5 17 June 2002 (‘UNESCO Final Report’): see especially at paras 81–100 and 112.
Cruelly, however, it stipulates that legislative ‘measures [shall] not, as a consequence, lead to the maintenance of separate rights for different racial groups and that they shall not be continued after the objectives for which they were taken have been achieved.’

In its 2006 response to the report submitted by South Africa under ICERD, the Committee singled out the provision in order to express its concern that our country’s affirmative action measures might indeed lead to the ‘maintenance of unequal or separate rights for those groups after the objectives for which they were taken have been achieved.’

Nothing in international law controverts ICERD – the Convention is exhaustive of the international position. The international position is clear: affirmative action, properly conceived, is legitimate and indeed mandated, but the division of people by race in order to subject them to segregated governance is not. International law, far from mandating multiracialism, positively reprehends it. Exponents of multiracialism, in whatever guise it is framed, will find naught for their comfort here.

V DOES THE CONSTITUTION SANCTION MULTIRACIALISM?

At the start of this essay, I pointed out, in general terms, that our Constitution espouses non-racialism. The provisions in question deserve a closer look. The Constitution opens by stating that the ‘Republic of South Africa is one, sovereign, democratic state founded on human dignity, the achievement of equality and the advancement of human rights and freedoms. Explicitly incorporated in this introductory clause, which is ‘foundational’ in its effect, is a commitment to ‘non-racialism and non-sexism’. After a few preliminary sections, a Bill of Rights is included that is far-reaching in scope and content. Drawing its inspiration from principles of constitutional supremacy envisioned by the founding fathers of the USA, the Constitution makes the Bill of Rights a charter against which all law, primary and subordinate, can be tested for coherence and validity.

Included in the Bill is the Equality Clause, which is to be found in s 9. Tracking its predecessor, which was in s 8, the clause, reads as follows:

1. Everyone is equal before the law and has the right to equal protection and benefit of the law.
2. Equality includes the full and equal enjoyment of all rights and freedoms. To promote the achievement of equality, legislative and other measures designed to protect or advance persons or categories of persons, disadvantaged by unfair discrimination may be taken.
3. 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.
4. 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.
5. Discrimination on one or more of the grounds listed in subsection (3) is unfair unless it is established that the discrimination is fair.

Emphasis supplied.

Concluding Observations of the Committee on the Elimination of All Forms of Racial Discrimination, South Africa, UN Doc. CERD/C/ZAF/CO3 of 19 October 2006, para 10.

See, for instance, Bato Star Fishing (Pty) Ltd v Minister of Environmental Affairs and Tourism & Others [2004] ZACC 15, 2004 (4) SA 490 (CC), 2004 (7) BCLR 687 (CC) para [73].

Section 1.
Its construction is framed by the interpretative rules embodied in the Constitution itself, and the same is true of enactments promulgated pursuant to it. They are three in number. The first is that the Bill of Rights must be construed in a manner that promotes the values underlying an open and democratic society based on human dignity, equality and freedom. The second is that parliamentary statutes and other enactments must be construed in a manner that promotes the spirit, purport and the objects of the Bill of Rights. The third is that international law ‘must’ and foreign law ‘may’ be considered when construing the Bill of Rights.

A The protection against arbitrariness: s 9(1)

Section 8 in the Interim Constitution opened by stating that ‘[e]very person shall have the right to equality before the law and to equal protection of the law.’ When s 9 replaced it, the right to ‘equal benefit of the law’ was added so that, under the present provision, ‘everyone is equal before the law and has the right to equal protection and benefit of the law.’

On the face of it, these rights are far-reaching, so much so that an ordinary person – one untutored in the law – might be forgiven for concluding that, since equality under law is ultimately achieved only when everyone is treated the same, the law cannot permissibly

---

44 The common law requires fealty to such directives, but the status of the Constitution as supreme law places the matter beyond doubt.


46 Section 39(1)(b) and (c).

47 Section 8 reads as follows: '(1) Every person shall have the right to equality before the law and to equal protection of the law. (2) No person shall be unfairly discriminated against, directly or indirectly, and, without derogating from the generality of this provision, on one or more of the following grounds in particular: race, gender, sex, ethnic or social origin, colour, sexual orientation, age, disability, religion, conscience, belief, culture or language. (3)(a) This section shall not preclude measures designed to achieve the adequate protection and advancement of persons or groups or categories of persons disadvantaged by unfair discrimination, in order to enable their full and equal enjoyment of all rights and freedoms (b) Every person or community dispossessed of rights in land before the commencement of this Constitution under any law which would have been inconsistent with subsection (2) had that subsection been in operation at the time of the dispossession, shall be entitled to claim restitution of such rights subject to and in accordance with sections 121, 122 and 123. (4) Prima facie proof of discrimination on any of the grounds specified in subsection (2) shall be presumed to be sufficient proof of unfair discrimination as contemplated in that subsection, until the contrary is established.’ The clause in the Interim Constitution, s 8, was substantially the same as the current one, s 9, so there is every reason to treat the judicial exegesis of the first as applicable to the second. The Court has certainly followed this approach. See Jooste v Score Supermarket Trading (Pty) Ltd (Minister of Labour intervening) [1998] ZACC 18, 1999 (2) SA 1, 1999 (2) BCLR 139 (CC) and cf National Coalition for Gay and Lesbian Equality & Another v Minister of Justice & Others [1998] ZACC 15, 1999 (1) SA 6, 1998 (12) BCLR 1517 (CC) para [15].
distinguish between its subjects and by so doing treat them differently. However, a moment’s reflection is all that is required to show that this cannot be the meaning of the section; if it were, the law would grind to a halt, since it typically operates, and must operate, by differentiating between subjects. If each and every differentiation made in terms of the law amounted to unequal treatment that had to be justified by means of resort to s 33, or else constituted discrimination which had to be shown not to be unfair, the courts could be called on to review the justifiability or fairness of just about the whole legislative programme and almost all executive conduct … The courts would be compelled to review the reasonableness or the fairness of every classification of rights, duties, privileges, immunities, benefits or disadvantages flowing from any law. Since this result would be intolerable, there must be a limit on what is constitutionally reviewable under the rubric of equality.

The limit resides, we can safely conclude, in the three phrases so far identified: ‘equality before the law’, ‘equal protection of the law’ and ‘equal benefit of the law’? In our case law, unhappily, no serious effort has been made to explicate them and delineate their scope. Our judges have been happy to speak about them in lofty terms. In Fraser v Children’s Court, Pretoria North, & Others, for instance, the majority proclaimed ‘that the guarantee of equality lies at the very heart of the Constitution [and] permeates and defines the very ethos on which the Constitution is premised.’ They have, in the same spirit, been happy to resort to comforting generalities. In City Council of Pretoria v Walker, the equal protection and benefit of the law was said to mean that ‘no one is above or beneath the law and … all persons are subject to law impartially applied’; and in National Coalition for Gay and Lesbian Equality and Another v Minister of Justice and Others, we learn that the equal benefit of the law means ‘that both in

48 This postulates, of course, equality between subject and comparator in every respect. From the discussion of Westen’s work, it should be obvious that such a universal identity between the two is seldom, if ever, comprehended by equality analysis. Typically, the comparison is of only some qualities and the point of reference is not co-extensive. Without doing violation to the principle of equality, I can say that this bag of beans is equal to that bag of sand by weight or in size. It is by reason of this fact that it is possible for a judge to say: ‘Equality means equal concern and respect across difference. It does not presuppose the elimination or suppression of difference. Respect for human rights requires the affirmation of self, not the denial of self. Equality therefore does not imply a levelling or homogenisation of behaviour or extolling one form as supreme, and another as inferior, but an acknowledgement and acceptance of difference.’ See Minister of Home Affairs & Another v Fourie & Another [2005] ZACC 19, 2006 (3) BCLR 355, 2006 (1) SA 524 (CC) para [60]. For dicta in the same vein, see the eloquent judgment of Sachs J in National Coalition for Gay and Lesbian Equality &Another v Minister of Justice & Others [1998] ZACC 15, 1999 (1) SA 6, 1998 (12) BCLR 1517 (CC) para [132].

49 ‘It must be accepted that, in order to govern a modern country efficiently and to harmonise the interests of all its people for the common good, it is essential to regulate the affairs of its inhabitants extensively. It is impossible to do so without differentiation and without classifications which treat people differently and which impact on people differently.’ Prinsloo v Van der Linde &Another [1997] ZACC 5, 1997 (6) BCLR 759, 1997 (3) SA 1012 (CC) (‘Prinsloo’) para 24.

50 Prinsloo ibid at para 17.


conferring benefits on persons and by imposing restraints on state and other action, the state had to do so in a way which results in the equal treatment of all persons. Yet, when the time came to give concrete effect to the clause, the Court treated it as superfluous by holding that it simply replicated the common law power of rationality review. The clause, it said, means no more than that a legal measure must have a legitimate purpose and make no distinctions between people that fail rationally to achieve the purpose.

For present purposes, therefore, the opening paragraph of the Equality Clause has nothing to tell us. Most certainly it cannot be read as mandating or legitimating multiracialism. When, as we do, we ask ourselves why the Court sanctions multiracialism, we will find no answer here. Such answer as emerges actually points the other way: non-racialism, not multiracialism, is the aim.

B The protection against unfair discrimination: s 9(3)

In Harksen v Lane, the Court revisited Prinsloo and reaffirmed its central message. Under ss (1), it said, an enactment that differentiates between people or categories of people will be condemned if it shows no ‘rational connection between the differentiation … and the legitimate governmental purpose it is designed to further or achieve’. However, the converse is not necessarily true: an enactment, though rational, can yet be impugned on the grounds that it constitutes unfair discrimination. ‘The determination as to whether differentiation amounts to unfair discrimination … require a two-stage analysis. Firstly, the question arises whether the differentiation amounts to “discrimination” and, if it does, whether secondly, it amounts to “unfair discrimination.” It is as well, the Court cautions us, to keep these two stages separate.

54 Ibid at para 27.
55 ‘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” 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 as well as to enhance the coherence and integrity of legislation. … Accordingly, before it can be said that mere differentiation infringes [subs (1)], 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. In the absence of such rational relationship the differentiation would infringe [subs (1)].’ Prinsloo (note 49 above) at paras 25–26. Generally see M Bishop ‘Rationality is Dead! Long Live Rationality! Saving Rational Basis Review’ (2010) 25 South African Public Law 312; A Price ‘The Evolution of the Rule of Law’ (2013) 130 South African Law Journal 649.

56 So much is all but explicit in the following dicta. ‘[W]hile the existence of such a rational relationship is a necessary condition for the differentiation not to infringe [the Equality Clause], it is not a sufficient condition; for the differentiation might still constitute unfair discrimination if that further element, referred to above, is present. … It is to [these provisions] that that one must look in order to determine what this further element is.’ Prinsloo (note 49 above) at paras 24 and 27. Unfairness in the course of discriminating is, it held, the further element.

58 Ibid at para 42 read with para 44. See too Prinsloo (note 49 above) at para 25. ‘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” 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 passage was cited with approval in City Council of Pretoria v Walker [1998] ZACC 1, 1998 (2) SA 363, 1998 (3) BCLR 257 (CC) para 27.
59 Harksen (note 57 above) at para 44.
60 Ibid at para 45.
Developing the theme, the Court explained that discrimination arises in circumstances in which there is unequal treatment of people based on attributes and characteristics viscerally inhering in them.\(^61\) The word ‘discrimination’ operates, in consequence, to limit the scope of the protection and so must be distinguished from mere differentiation. Enactments that differentiate between persons or classes of person are in themselves unobjectionable, but they become discriminatory if they have the potential to affect the comparative wellbeing of persons or groups within the state in a way that, for present purposes, we can describe as visceral. The impact must, the judge continues, be deleterious. ‘Given the history of this country, “discrimination” has acquired a particular pejorative meaning relating to the unequal treatment of people based on attributes and characteristics attaching to them.’\(^62\) This requirement, said the Court, is deemed to be satisfied if the distinction is made by reference to one of the factors, of which race is but one of more than a dozen,\(^63\) specifically listed in the section. Otherwise, its existence depends on ‘whether, objectively, the ground [of differentiation] is based on attributes and characteristics which have the potential to impair the fundamental dignity of persons as human beings or to affect them adversely in a comparatively serious manner.’\(^64\)

Section 9(3), in summary, provides a wholehearted endorsement of non-racialism. To say this is not to say that it outlaws all distinctions based on race. Some can be imagined that a court would treat as legitimate under the section: selecting a black person to play Othello, perhaps, or to sell goods door-to-door to a customers who, being black themselves, tend to prefer dealing with blacks. On top of this, the clause must be placed in the context of a section that plainly creates an exception – or, if you prefer, qualification – in favour of measures designed to redress past discrimination.\(^65\) Making due allowance for such reservations, however, it seems fair to say that, if the foundational section (s 1) sets up non-racialism as a goal of our society, this is the clause that, by prohibiting discrimination of an unfair sort, creates the instrument by which the goal can be realised. Taken in the round, it is a broad negation of race as a proper basis of social governance and operates, in consequence, against the notion that multiracialism is a competent policy for the purpose.

\[\text{C} \quad \text{The recognition of restorative measures to promote equality: s 9(2)}\]

The Equality Clause has two basic elements. One comprises the prohibition with which we have just been dealing, namely, that there shall be no unfair discrimination. It is negative in nature – it prohibits. The other comprises a right that is positive. It says that people are entitled to equality and, if they have been denied this right by past unfair discrimination, measures to repair the harm are legitimate, indeed desirable. The right is created by s 9(2), which states as follows: ‘Equality includes the full and equal enjoyment of all rights and freedoms. To promote

\(^{61}\) Ibid at para 46.

\(^{62}\) Ibid at para 46 citing Prinsloo (note 49 above) at para 31.

\(^{63}\) Fourteen then, sixteen now: ‘race, gender, sex, pregnancy, marital status, ethnic or social origin, colour, sexual orientation, age, disability, religion, conscience, belief, culture, language and birth.’

\(^{64}\) Ibid at para 52)[(b)(i)]

\(^{65}\) Sensibly, the Court gives effect to the distinction, a necessary one, by treating dignity as the fulcrum upon which s 9(3) pivots. Differentiation that impairs dignity is treated as proscribed under the section.
the achievement of equality, legislative and other measures designed to protect or advance persons or categories of persons, disadvantaged by unfair discrimination may be taken. 66

The text of the clause, which is enigmatic, presents two problems. The word ‘includes’ in the phrase ‘equality includes the full enjoyment’ gives rise to the first. A provision framed in such language, the language of ‘inclusion’, plainly posits a generic set of rights within which the specific, designated, rights are included. The first, and perhaps more linguistically natural, construction of this language is to treat the sentence as an elaboration of the rights just delineated: read thus, the sentence is conceived as relating back to ss (1), and serves to locate the specified rights and freedoms among those referred to in the earlier provision. The second posits no obvious connection between the two clauses and, by making ss (2) self-standing, a set of rights and freedoms, independently sourced, of which the full and equal enjoyment of all rights and freedoms are but some. Textually, the second construction makes better sense, not least because ss (1), referring only to rights and making no reference to freedoms, can provide no foundation, at least literally, for a construction sufficiently over-arching to embrace the ‘freedoms’ referred to in ss (2). Practically, however, it seems to make no sense. The expression ‘the full and equal enjoyment of all rights and freedoms’ is so broad as to appear exhaustive, and it is hard to imagine anything falling beyond its compass.

There is, I venture to suggest, no getting to the bottom of this question and I doubt whether much turns on it. It is the spirit of the provision, rather than the words it employs, that principally seems to matter. What we learn from it, basically, is that people are entitled to enjoy their rights and enjoy their freedoms whomsoever they may be. This may seem glib, but so be it. Such cavalier handling of the second problem in the section cannot so lightly be countenanced, however. It arises out of the juxtaposition in the sub-section of the two ideas separately embodied in the two discrete sentences. The first sentence speaks of equality in terms that are general and very wide. It says that equality must be taken to include the ‘full and equal enjoyment of all rights and freedoms’, but what it does not identify is the subject, the repository, of the rights and freedoms. Is it the individual or is it the collective? More pertinently, if it denotes the collective, does it comprehend collectives defined by race and disposed of accordingly? If so, it might, were it taken in isolation, provide some basis, however tenuous, on which to construct a system of multiracialism. Whites, coloureds, Indians and blacks, when so categorised, must each receive their ‘full rights and freedoms’ and this is attainable by recognizing the racial categories and separately endowing their members with the benefits in question.

As I say, this construction might be tenable if the sentence operated in isolation, but of course, it does not – it is connected to a subsequent section that, in terms, renounces discrimination. By coupling the two sentences together, the lawgiver must, on basic principles of construction, intend that each should be evocative of the meaning of the other. In specifically legitimizing measures designed to promote the interests of persons previously disadvantaged by unfair discrimination, the second sentence seems plainly intended to sharpen the focus and narrow the scope of the antecedent sentence. So much would be the effect of an application of

66 In s 8 of the Interim Constitution, the equivalent provision (ss 3(a)), took the form of a savings clause – the section is not to be read as ‘precluding’ measures taken for the specified purpose. The new section reads better, but the change is not purely ornamental. The underlying object is to emphasise that the measures in question constitute no exception to the tenet of equality but are an elaboration of it. They are designed to entrench the constitutional legitimacy of substantive equality and so disavow a construction, which was tenable under s 8, that the provision denotes only formal equality.
the *eiusdem generis* rule and it seems peculiarly apt in the context.\(^{67}\) If this is correct, then the section, read as whole, must be taken to comprehend only measures that are restorative, that is, that are designed and tailored to remedy the consequences of past unfair discrimination.

This conclusion is in no way undermined by the inclusion in the clause of a reference to ‘categories of persons’ disadvantaged by past unfair discrimination. Underlying this wording, firstly, is a desire to ensure that a measure will survive scrutiny despite the fact that some who profit from it have not personally suffered unfair discrimination. Moreover, extending the scope of the clause to cover groups reflects an understanding of how hard it can be to decide the extent to which discrimination, as opposed to other causes, is responsible for a state of deprivation. Problems of causation such as this pervade the law and can only be resolved on a common sense basis.\(^{68}\)

Section 9(2), we conclude, provides no support for multiracialism. It contemplates a world in which South Africans will enjoy equal rights and freedoms and, for this purpose, legitimates the systematic reversal of the consequences, to individuals and groups, of past discrimination, race-based discrimination for present purposes. Multiracialism, which looks to a socially constructed future, is wholly unconcerned with past discrimination. It defines people, both individuals and groups, by race and, whenever it discovers the existence of some differential, it devises some remedy that will eliminate the disparity. It is unconcerned with the source of the difference – whether it be self-created, innate or externally generated – and equally unconcerned with issues of fair-play, morality or comparable justification. Whites will be remediated if they are the underdogs in just the same way as blacks might be under the same circumstances. Placing a construction on s 9(2) that sanctions such a programme is wholly foreign to the text once one appreciates that the second half of the sub-section influences and shapes the way the first half is construed. Taken in isolation, the first sentence of s 9(2) might conceivably be read as countenancing the separate handling of the races, but when the two sentences are read together, no such interpretation is tenable.

Nothing, therefore, in s 9(2) can be read as legitimizing multiracialism. The same is true, we can accept, of the balance of the Equality Clause and indeed the Constitution at large. There are, to be sure, a few clauses in the document that require state institutions to be broadly representative of the people of South Africa,\(^{69}\) Dealing with them in detail is beyond the scope of this article; here it is enough to say that efforts to build a justification for multiracialism on the basis of them are bound to prove futile. They are, I suggest, designed to do no more than prevent parochialism, by which I mean, an impulse to favour one’s own group over others. Certainly they cannot be construed as legitimizing a structure of race-based streaming wholly anathema to the concept of non-racialism.

---

\(^{67}\) The rule is but a guide. I say its application is apt not just because of the juxtaposition of the two sentences, which suggests that the two sentences are interrelated, but also because the repetition of the word ‘equality’ in the opening words of the second sentence cannot but be a reference back to the ‘equality’ that opens the first sentence.

\(^{68}\) Revealing an understanding of this fact, Justice Mokgoro put the issue on the footing that there need not be ‘a rigid link between the nature of the disadvantage suffered … and measures taken to alleviate that disadvantage’ [but at least] ‘there should be some correlation between the two.’ *President of the Republic of South Africa & Another v Hugo* [1997] ZACC 4, 1997 (6) BCLR 708, 1997 (4) SA 1 (CC) para 94.

\(^{69}\) See Constitution, ss 174 and 195.
VI TOO BAD, THOUGH: THE COURT THINKS MULTIRACIALISM IS JUST FINE

In the face of these provisions and of the context in which they are set, the Court has given its imprimatur, expressly and unreservedly, to multiracialism. Let us have no doubt about what this means. Seduced by the siren-song of demographic representivity, the Court has held that the people of South Africa can justifiably be (1) separated into racial categories (2) comprising the matrix of white, coloured, Indian, and black created by the apartheid regime (3) and then be managed, controlled, favoured and disfavoured according to the category into which they fall (4) provided only that the groups are not disproportionately treated.

What paved the way for this result was Van Heerden and, though it was not directly concerned with multiracialism, it must be carefully considered for the latitude it created. Barnard tripped lightly down the path, to be followed by Correctional Services until Insolvency Practitioners signalled that the destination had been reached, and each of them deserve a corresponding level of scrutiny. The path, it will be seen, has been a slippery one.

A Van Heerden

In the course of negotiations that ushered in our democracy, it was decided that a special fund known as the Central Provident Fund (‘CPF’) should be created to enhance and ring-fence the pension entitlements of old-order members of the Parliament. Some time after the election of the new order Parliament, provision was made by statute for a new fund that gave a temporary preference to members elected for the first time to the new Parliament. The essence of the claim in the case, Minister of Finance & Other v van Heerden, was that the new scheme ‘improperly disfavours … members who are in receipt of pensions from the CPF in comparison with new parliamentarians who … do not receive pension benefits from the CPF.’ The impropriety was said to consist in an intent to prefer members ‘based on intersecting grounds of race and political affiliation.’ The relief claimed was an order entitling the CPF members to the same benefits under the new fund as those given to the newcomers to the new Parliament.

In the High Court, the measures were held to be discriminatory in their effect and to require proper justification before they could be treated as legitimate. In the ensuing appeal, Moseneke J, writing for the overwhelming majority of the judges, rejected this approach as quite wrong. He began by proclaiming, uncontroversially, that –

[r]emedial measures are not a derogation from, but a substantive and composite part of, the equality protection envisaged by the provisions of section 9 and of the Constitution as a whole. Their primary object is to promote the achievement of equality. To that end, differentiation aimed at protecting or advancing persons disadvantaged by unfair discrimination is warranted provided the measures are shown to conform to the internal test set by section 9(2).
In dealing with the issue, the court below had relied on the reverse onus provision in the section (s 9(5)), but this was not designed with cases such as this in mind, in which the object of the enactment was ameliorative. Moseneke J stated as follows:

‘It seems to me plain that if restitutionary measures, even based on any of the grounds of discrimination listed in section 9(3), pass muster under section 9(2), they cannot be presumed to be unfairly discriminatory. To hold otherwise would mean that the scheme under s 9(2) is internally inconsistent or that the provisions of section 9(2) are a mere interpretative aid or even surplusage. I cannot accept that our Constitution at once authorises measures aimed at redress of past inequality and disadvantage but also labels them as presumptively unfair. Such an approach, at the outset, tags s 9(2) measures as a suspect category that may be permissible only if shown not to discriminate unfairly.’

The reverse onus clause should not be made to govern the affirmative action provision. Were it to do so, the effect would be intolerable, since ‘presumptive unfairness would unduly require the judiciary to second-guess the Legislature and the Executive concerning the appropriate measures to overcome the effect of unfair discrimination.’

Ordinary discrimination cases – that is, cases in which no remedial object is being pursued – should, in his mind, continue to be assessed by reference to fairness, but measures whose ostensible object was restorative should be subject to less strenuous scrutiny. ‘When a measure is challenged as violating the equality provision, its defender may meet the claim by showing that the measure is contemplated by s 9(2) in that it promotes the achievement of equality and is designed to protect and advance persons disadvantaged by unfair discrimination.’

If the motive was remedial, the person challenging the measure would have to discharge the onus of establishing that the means being employed for the purpose were irrational. The proper approach to such cases, said the Justice, was to posit three requirements:

The first yardstick relates to whether the measure targets persons or categories of persons who have been disadvantaged by unfair discrimination; the second is whether the measure is designed to protect or advance such persons or categories of persons; and the third requirement is whether the measure promotes the achievement of equality.

---

76 The reasoning is, to put it mildly, unpersuasive. A reverse onus clause operates purely within the domain of procedure: it simply determines who bears the burden of persuasion and, in consequence, serves only to identify who will lose if the evidence is unconvincing. In its conclusion the judgment is equally unpersuasive. In law, the legitimacy of the conduct of a repository of power is seldom made to depend on the presence or absence of proper motive. Acts are typically struck down even when the motive is benign and upheld even when malign. Issues of remedial justice for past discrimination are, by international law, exemplars of this principle. So much is clear from the comprehensive report of Rapporteur Bossuyt in the UNESCO Final Report (note 39 above). Over and over, he emphasises that motive should not provide a basis for the justification of a remedial measure. See at paras 107, 108 and especially at 113, where he says: ‘The aim or goal is not decisive …. Affirmative action should not be interpreted as justifying any distinction based on any ground with respect to any right merely because the object of the distinction is to improve the situation of the disadvantaged individuals or groups. Affirmative action is no exception to the principle of non-discrimination. Rather, it is the principle of non-discrimination that establishes limits to each affirmative action.’

77 The argument is, of course, conceptually untenable. The core function of the Court is to ‘second-guess’ the Legislature and the Executive by considering, in the course of litigation, whether their enactments comply with the imperatives of the Constitution. The issue is not whether to second-guess, but in what manner and to what degree. Second-guessing is, to be sure, impermissible if what is being second-guessed is the Constitution itself. This is, in effect, what the Court did here.

78 Van Heerden (note 70 above) at para 33.

79 Ibid at note 37.
If all three requirements were met, the measure would be unimpugnable.

The language in which the majority’s test is framed may be oblique but its object is plain. The motive behind the measure provides the criterion that distinguishes one form of assessment from the other. Race-based measures are prohibited if they are unfairly discriminatory, but measures designed to favour people previously disadvantaged will survive scrutiny if they are ‘reasonably capable of attaining the desired outcome’. A measure will have the character of rationality if it satisfies the test by which provisions are evaluated for rationality under ss (1). ‘If the remedial measures are arbitrary, capricious or display naked preference they could hardly be said to be designed to achieve the constitutionally authorised end.’ Moreover, said the Justice, deference is the court’s proper response in the face of a challenge to an equalizing measure; the courts, in his view, must give decision-makers wide scope for the implementation of such measures. There is no place for the ‘tightly circumscribed affirmative action’ that would otherwise pertain.

Reasoning thus, the Court achieved two things: it relaxed the test by which measures ostensibly intended as remedial would henceforth be evaluated and placed the onus of proof back on the party mounting the challenge. Unlike ordinary measures of a discriminatory nature, which would be struck down if unfair, those designed to be remedial would survive scrutiny provided rational means were employed to achieve their end. Proof of the requisite objective would, it seems, remain the duty of the ‘sponsor’ of the measure, but little if anything would be required beyond a showing that the measure’s beneficiaries were people who had previously been the victims of discrimination. Once this onus was discharged, the challenger would be saddled with the burden of proving that the means employed in the measure bore no rational connection to the objective sought to be attained.

By reasoning in this way, the Court placed remedial measures beyond the reach of anti-discriminatory provisions of the Constitution, and so put their evaluation on the same footing as ordinary enactments. Challenges might still be mounted on the conventional grounds of constitutional review, but no special protections against discrimination were to be derived from s 9. A measure intended to favour blacks, presumptively regarded as the victims of past discrimination, would survive scrutiny provided it promoted their interests in a rational way. It would, in effect, be beyond the purview of s 9. In contrast, a race-based measure revealing no such intention would be subject to scrutiny under the section and survive only if it met the standard of fairness posited by the concept of unfair discrimination.

By its very nature, a measure promoting multiracialism can have no such intention. Its object is to stream people by race and treat them commensurately within the streams so created. Its intention, taken on its face, cannot be intended to favour one race above another. As a result, it should be subjected to scrutiny by reference to the stricter standard of fairness and be treated as constitutional only if, in accordance with the reverse onus requirement in sub-clause (5), the ‘sponsor’ of the scheme can show that it passes muster as fair. So much should be obvious, but would the courts treat it as such?

80 Ibid at para 41.
81 Ibid.
82 Ibid at para 2.
83 There is some debate over whether the Judge intended so parsimonious a test, but this is how I read the passage and there is nothing in the ambient jurisprudence to say that my interpretation is wrong.
84 In its application it may have this effect, of course. It will do so when one race – black or white, it makes no difference – is under-represented in the class of persons under consideration.
In recent years they have given their answer to this question in three cases: *Barnard*, *Correctional Services*, and *Insolvency Practitioners*. They are considered below and, properly understood, they deliver an unequivocal answer. In summary, it is No – multiracial measures will not be tested by reference to fairness and the reverse onus will not be applicable in considering them. The rationality criterion will govern them and the challenger, not the ‘sponsor’, will bear the onus of proving irrationality.

**B Barnard**

In *South African Police Service v Solidarity obo Barnard*, it was common cause that the application for promotion by Captain Barnard, a white, female police officer, was twice recommended by a multiracial committee as the best candidate on merit. However, each time she found herself knocked back on the grounds that, on a grid that divides personnel by race and gender, people with her attributes – that is, white women – were over-represented in the grade when seen as a proportion to their numbers in the population as a whole. On each occasion, the post was left unfilled so Barnard continued to perform the role in an acting capacity.

The argument mounted on behalf of Captain Barnard, reduced to its essence, was that the use of such a grid to stream applicants by race-cum-gender can axiomatically never constitute a measure whose object is to rectify past discrimination. The reason becomes plain once it was understood that, under such a scheme, a white male and so, by definition, a person advantaged by past discrimination will be preferred, whatever his relative merits, if every other race-cum-gender group is over-represented in the grade in question. The scheme was a multiracial one, pure and simple. Being so, its legitimacy had to be determined by reference to the anti-discrimination provisions in s 9 of the Act, which had received their expression in the comparable provisions of the Employment Equity Act 55 of 1998.

Justice Moseneke, again writing for the majority, opened his judgment by trumpeting the virtues of non-racialism. ‘Our constitutional democracy is founded on explicit values. Chief of these, for present purposes, is human dignity and the achievement of equality in a non-racial, non-sexist society under the rule of law.’ If we are to achieve this aim, he explained,


86 Grids setting race-cum-gender standards by which the staff complement is to be managed are now commonplace in both the private and the public sector. The Plans in which the grids are to be found generally follow a pattern. The staff complement of the employer is distributed in each grade by gender (male and female) and race (white, coloured, Indian or African). The eight subsets in a grade (four races multiplied by two genders) are then compiled into a grid. Each number is then reduced to a percentage of the total staff component within a grade and this is then compared to the equivalent demography, by race and gender, of the population of the country as a whole (or, sometimes, the economically active subset of it.) By these means the employer can establish whether a particular category of person within a grade is ‘over-represented’ or ‘under-represented’ and accept or reject applications for recruitment or employment accordingly. If the post cannot be filled because the only suitable candidate is ‘over-represented’, the post will be left vacant unless the employer overrides the grid in order to secure special skills or to satisfy some other exiguous condition.

87 After the case, she sought and obtained employment with Solidarity, her trade union. What does not emerge from the case, but is worth noting, is that her father was himself a police officer and had brought her up to believe that serving in the Force was an honour of the highest sort.

88 Section 10.

89 *Barnard* (note 85 above) at para 28.
remedial measures must be implemented, but they must remain ‘within the discipline of the Constitution [so that they do not] unduly invade the human dignity of those affected by them, if we are truly to achieve a non-racial, non-sexist and socially inclusive society.’ It is only by these respecting this mandate that we can achieve the ultimate goal of ‘a more equal and fair society that hopefully is non-racial, not sexist and socially inclusive.’

After so promising a start, the judge might have been expected to examine the Department’s scheme in order to determine whether it constituted a remedial measure and, if it did, whether it satisfied the requirements of non-racialism and non-sexism. If he had, the conclusion would have been inescapable: the scheme, by treating white males in the same way as everyone else, cannot conceivably be remedial. Once properly examined, its true object became plain. The Department was dividing people by race and gender and using the resulting grid to secure proportional demographic representation within the workplace. Such a scheme, classically one of a multiracial nature, constituted naked social engineering far removed from a properly-conceived system, that is, one ‘directed at remedying past discrimination … formulated with due care not to invade the dignity of all concerned.’ The application of the Plan was unconstitutional and a proper remedy had to be fashioned for Captain Barnard, who had been wronged by reason of its implementation.

This is not how the learned judge reasoned, however. Woefully failing to engage with the central issue in the case, he took it as read that the Plan constituted a measure designed to remedy past discrimination even though, on its own terms, it was characterised as a mechanism by which demographic representation within the service would be secured. Then he recited the Van Heerden test even though, once again expressly, its scope is limited to measures that ‘target a particular class of people who have been susceptible to unfair discrimination’.

The test, he complacently recorded, gave the Court only a limited capacity to review the decision to refuse Barnard’s promotion. It entailed an application of the principle of legality which meant that the decision, to be ‘a legitimate restitution measure[,] must be rationally related to the terms

90 Ibid paras 28 and 32 respectively.
91 Ibid at para 30.
92 Barnard (note 85 above) at para 30.
93 The fact that no challenge had been mounted against the Plan itself was not regarded by the judge as an obstacle to relief. Even presuming the Plan to be valid, there ‘is no valid reason why courts are precluded from deciding whether a valid [Plan] has been put into practice lawfully.’ Ibid at para 38.
94 For the sake of clarity, I have concentrated on the cardinal jurisprudential issue, but there are two points, peculiar to the case, that I gloss over in the text. It is as well that I deal with them, albeit briefly, before proceeding. Both stem from the fact that the Plan itself was never subjected to formal challenge in the case.
1. The first is the contention that, in the absence of a challenge to the Plan formally mounted by way of proceedings on review, a decision made pursuant to it had to fail. (Ibid at para 60.) This conclusion is wrong in law. Plans are framed and filed by the state in its capacity as an employer, and such conduct (not being state action) is not susceptible to review: Chirwa v Transnet Limited & Others [2007] ZACC 23, 2008 (4) SA 367, 2008 (3) BCLR 251, [2008] 2 BLLR 97, (2008) 29 ILJ 73 (CC).
2. The second is that ‘Mrs Barnard accepted that the Plan was a valid affirmative action measure’ (ibid at para 52). There is simply no basis for this finding, and it is wholly misconceived. The case mounted on her behalf accepted that the legitimacy of the Plan had not formally been challenged but proceeded on the basis that, since it was bad in law, its implementation was equally bad. The concession concerned a matter of procedure; the argument was based on the substance of the measure.
95 Barnard (note 85 above) at para 36.
and objects of the measure." The onus to show this was hers to discharge. On the facts, she had failed to do so.

In their minority judgment, Cameron, Froneman and Majiedt JJ took a different view. They held that there is 'a tension between the equality entitlement of the individual and the equality of society as a whole.' Resolving the clash requires judicious decision-making. A balance had to be struck that, they implied, goes beyond the mechanical application of a formula. Fairness should be the touchstone. Van der Westhuizen J, for his part, stressed that the interests of the group cannot be determinative since measures to promote equality can affect individual dignity. A balance, he emphasised, had to be struck between the two constitutional imperatives in order to resolve the tension immanent within them.

Reading these minority judgements in the round, it seems permissible to conclude that the judges deprecated the use of grids and the multiracialism they embody. Justice Jafta had no such qualms, but placed himself firmly in the camp of multiracialism. 'By not appointing Ms Barnard and reserving the post for black officers, the National Commissioner sought to achieve representivity and equity in the Police Service. This accords with the Employment Equity Plan and is consistent with the purposes of the Act. Therefore, the National Commissioner’s decision cannot constitute unfair discrimination nor can it be taken to be unfair.' The reverse is true, however. Representivity, pursued as a goal, is nothing but multiracialism dressed up in different clothes.

C  Correctional Services

_Solidarity & Others v Department of Correctional Services & Others_ was a similar kind of case. The action was brought by eleven prison warders in order to obtain the promotion they had been refused because, given their colour and gender, they were demographically 'over-represented' in the categories of posts to which they sought elevation. In contrast to _Barnard_, they happened to be 'coloured' in the main and by no means were all of them male.

As in _Barnard_, the contention was that the Plan was not designed to redress past discrimination but embodied a form of social engineering that was impermissible under the Constitution. In _Barnard_ the contention had found no favour with the majority, who held that the Commissioner had properly ‘exercised his discretion not to appoint Ms Barnard, even though she had obtained the highest score, because her appointment would have worsened the representivity in [the target category]' .

Building on this statement, Zondo J, writing for the majority, embraced the concept of representivity in the following words:

---

96 Ibid at para 39.
97 Ibid at para 53.
98 Ibid at para 77.
99 On the facts, however, he felt that Barnard’s personal interest should bow before those of the group.
100 (Ibid at 227). Endorsing the decision of the Labour Appeal Court, he held that, ‘because “the essence of restitutionary measures is to guarantee the right to equality”, their implementation cannot be subject to an individual’s right to equality.’ (At para 231). The statement is a surprisingly blunt endorsement of the importance of multiracialism.
102 Ibid at para 62.
Black candidates, whether they are African people, Coloured people or Indian people are also subject to the *Barnard* principle. Indeed, both men and women are also subject to that principle. This has to be so because the transformation of the workplace entails … that the workplace should be broadly representative of the people of South Africa.\(^{103}\)

Then, to his intellectual credit, the judge recognised the implications of what *Barnard* portended and he was now saying. Representivity entailed the division of the workforce into streams based on race (plus gender) and decision-making that ensured the proportionate distribution of each segment in every layer of the workforce.

A workplace or workforce that is broadly representative of the people of South Africa cannot be achieved with an exclusively segmented workforce. For example, a workforce that consists of only White and Indian managers and, thus, excludes Coloured people and African people or a senior management that consists of African people and Coloured people only and excludes White people and Indian people.\(^{104}\)

In his view, it was only by effectuating such a distribution that equity could be achieved. Anything less would be ‘unacceptable’.

It would be unacceptable, for example, for a designated employer to have a workforce of five hundred employees fifty of whom occupy senior management positions but only five of those senior management positions are held by African people when twenty are held by White people, fifteen by Coloured people and ten by Indian people despite the fact that in the population of South Africa, African people are by far the majority.\(^{105}\)

The Employment Equity Act, which provided the immediate basis for the challenge, required a construction in conformity with the Constitution. Upon a proper interpretation, the Act must be seen as striving for ‘a constitutional objective that every workforce or workplace should be broadly representative of the people of South Africa. … It will not be enough to have one group or two groups only and to exclude another group or other groups on the basis that the high presence of one or two makes up for the absence or insignificant presence of another group or of the other groups.’\(^{106}\) A grid, race-based though it might be, was a proper basis upon which to determine matters of appointment or promotion, and all that mattered was whether the apportionment was rationally contrived.\(^{107}\) (In the instant case, so it happens, the apportionment was held to be irrational as no account had been taken of regional imbalances. For this reason, and for no other, the applicants succeeded in their challenge to the grid.)

Justice Nugent, with whose judgment Cameron J concurred, felt considerably less sanguine about the Plan. To him, it seemed coldly calculating.

There is no sign in the Plan now before us of the just balancing required by [earlier CC cases] nor is there any recognition of the care and vigilance expressed in *Barnard*. Nor is there any attempt to harmonise the constitutional tension that concerned the concurring judges, nor of the balancing urged by Van der Westhuizen J. In contrast to the thoughtful, empathetic and textured plan one might expect …what we have before us is only cold and impersonal arithmetic.\(^{108}\)

He accepted that ‘for powerful historical reasons the statute has focused on race and gender as

\(^{103}\) Ibid at para 40.

\(^{104}\) Ibid.

\(^{105}\) Ibid.

\(^{106}\) Ibid at para 49.

\(^{107}\) ‘(T)he determination that the group is already adequately represented or overrepresented [must have] a proper basis’ (ibid).

\(^{108}\) Ibid at para 102.
markers of employment equity’\textsuperscript{109} and that its purpose is the attainment of ‘representivity in the workplace,’\textsuperscript{110} but this could not be done in so mechanical a way.

That [mandated] goal is capable of being achieved only by a visionary and textured employment equity plan that incorporates mechanisms enabling thoughtful balance to be brought to a range of interests. … Ours are a vibrantly diversified people. It does the cause of transformation no good to render them as ciphers in an arid ratio having no normative content.\textsuperscript{111}

The judge was right. Gradgrind\textsuperscript{112} might be proud of such a system, but we should not be.

D \textbf{The Insolvency Case}

The remorseless logic of multiracialism reached its apex in \textit{Minister of Constitutional Development and Another v South African Restructuring and Insolvency Practitioners Association and Others}\textsuperscript{113} (the ‘Insolvency Case’). The factual matrix takes a form that by now will be familiar. Under prevailing insolvency legislation, the Master of the Court is entrusted with the power to appoint provisional liquidators to take charge of the estate in the period, which can be considerable, that elapses before a final liquidator is appointed. Over the years the Master developed a system under which creditors, whose interests are the only ones at stake, would file ‘requisitions’ proposing candidates for the appointment of the liquidator they preferred. In general, the system intended to endow those liquidators who, in the eyes of the creditors, could be expected to wind the estate up most effectively. Traditionally they have been overwhelmingly white and, for years, the imbalance has been keenly felt. Over the past two decades the Master, supported by the industry, has elected to supplement the appointees by adding a black person to the team who, by learning the ropes, might promote the cause of transformation in this field. Established liquidators, subscribing to the object, were content with the system even though a significant portion of their fee would now go to a person they regarded as supernumerary. Creditors, for their part, were unconcerned since only a single fee remained payable.

\textsuperscript{109} Ibid at para 121.

\textsuperscript{110} Ibid at para 125.

\textsuperscript{111} Ibid at para 133.

\textsuperscript{112} The reference is aptly described in Wikipedia. ‘Mr Thomas Gradgrind is the notorious school board Superintendent in Dickens’s novel \textit{Hard Times} [whose] name is now used generically to refer to someone who is hard and only concerned with cold facts and numbers.’

\textsuperscript{113} [2018] ZACC 20 (CC), 2018 (5) SA 349 (CC); 2018 (9) BCLR 1099 (CC).
The scheme has proved over time to be less than completely successful, and the Chief Master, encouraged by the Department, decided it needed an overhaul. What he crafted and promulgated was a proposal that stretched multiracialism to its limits. What he envisaged was a system in which appointments would be made strictly in proportion to the national demographics of race and gender. In terms of the scheme, insolvency practitioners had to be appointed in terms of a grid whose application would be mechanical and strict. While exceptions would be entertained in especially demanding cases, appointments would ordinarily be made consecutively in the ratio A4: B3: C2: D1, where –

- and the numbers 4: 3: 2: 1 represent the number of insolvency practitioners that must be appointed in that sequence in respect of each such category.

Justice Jafta, who penned the majority judgment, began by issuing a clarion call of the sort that customarily adorns judgments in this field. The evils of the past, the consequence of nakedly racist attitudes, had created imbalances that required remedial action and the policy with which the Court was now concerned was, in his view, just such a restitutionary measure. Counsel for the practitioners was wrong to suggest that the policy was not designed to redress past injustice but to operate as a system of race-based social engineering. In his view, the scheme was a proper mechanism of redress and this, it seems, was true even though whites would, in rotation, profit by it as well. He could see nothing wrong with a policy that, ignoring the merits of individual practitioners and the wishes of creditors, made race the sole determinant for appointment. Multiracialism, rigorously employed, was to be the touchtone of the system.

VII WHY IS THIS ALL WRONG?

Multiracialism, I have contended, is a bad thing. There is no need to take my word for this. Our common law, which is simply common sense writ large, reprobates the use of race-based classifications in the control and regulation of society. Courts steeped in the common law, generally reject them when they can, even if the implementation of the system of separation is grounded on or designed to encourage equality.

When mandated by a race that is socially superordinate, ‘separate but equal’ is a pernicious doctrine that serves to demean, marginalise and subjugate members of the subservient race. Brown tells us as much, and so do the minority judgments in Plessy and, closer to home, Rasool. Our lived experience tells us the same thing about the policy of separate development, a policy consciously crafted in the belief that races are to be separately defined, disaggregated and managed. Whether cynical or genuine in conception and implementation scarcely matters for present purposes: what does signify is that it perpetuated the subordination of black people and stripped them of their dignity as individuals. In dealing with matters of race, judgments

114 Knowing they would not ultimately be accountable, black supernumeraries often failed to pull their weight. They were happy to take the money but not to do the work. There is, I hazard, nothing surprising in this.

115 The disparate treatment of citizens by naturalization was the principal reason why the scheme was condemned as illegitimate by the Court.
repeatedly recall the horrors of our past and, however repetitious this can seem, the courts cannot be castigated for doing so. We need to be reminded, and reminded frequently, of our past lest, by forgetting, we perpetuate the cycle of race-based repression: here, as elsewhere, ‘never again’ is an excellent cri de coeur.

Does ‘separate but equal’ become acceptable if mandated by a race that is socially subordinate? The ready answer is yes: the policy becomes a means by which imbalances will be systematically redressed without traducing the principles of formal equality. When black people have been shut out of the workplace throughout many years of racial discrimination, they will naturally find themselves under-represented as a group, especially in the upper echelons of the hierarchy. If racial discrimination is exclusively responsible for the disparity, a good moral case exists for redressing the imbalance; if not, the programme can still be justified as a means by which black people can, in the interests of social cohesion, be ‘brought into the economy’. A policy designed to produce proportional representation by race will naturally have an equalizing effect and it will provide a plausible answer to whites who complain that they are now become the victims of reverse discrimination. So much certainly underlies the thinking of the Court in the judgments considered above and Justice Zondo must be thanked, for this if nothing else, for taking the trouble to make the attitude explicit.

These contentions contain the essence of the case for multiracialism. Before making an effort to reply to them, as obviously I must, I should make the battle-lines plain. Nowhere in the essay have I contended that race should never be a basis for decision-making by the state. Such an argument can, in the abstract, be made, but as a matter of law, it would be hard to sustain in the face of the express provisions of s 9. They seem plainly to permit, even mandate, a measure of race-based action, and have been authoritatively so construed by the Court. Placing the matter squarely within these parameters, it need hardly be said, reduces the scope of the issues. It all but eliminates moral and ethical questions – is it right to make race a criterion for decision-making? – and, in the same vein, it all but eliminates such mechanical and functional questions as whether there is truly such a thing as race and, if so, how do we decide into which racial category a person falls. By reducing the issues in this way, we discover that only one remains – ‘how much’? If we want a system of race-based classification in all its glory, then multiracialism should be our choice. If not, we should prefer the more modulated system that we have learned to call affirmative action, where race is but one factor in the plurality of circumstances by which the decision is informed.

The answer appears simple enough on the face of it. At the outset of this essay, we reminded ourselves that equality is an empty idea which we have to fill by reference to our societal values. We have come full circle, only now we must leave the realms of the abstract and fill the void concretely. Let us tap into our societal values, then, and see what we bring back.

In the absence of an obvious consensus about what these values might be, we, as lawyers, must seek them in the sources I have described above – the Constitution, first and foremost, and then common law. Taken in the round, they can, I maintain, be said to favour the individual over the collective, liberty over coercion and remedy over regulation. They countenance the employment of the race-based criterion to redress the harm caused to an individual by the discriminatory use of the self-same criterion. But in doing so, they invoke the criterion only when it is reasonably necessary to achieve the remedial purpose and anxiously guard against its gratuitous employment for goals going beyond that. They most certainly cannot be taken to espouse the broad-ranging social engineering, arbitrary in form and rigid in content that
finds its expression in the kind of multiracialism so favoured by judges such as Zondo J. They recognise, first, that streaming people in this way produces injustices to individuals such as Captain Barnard; secondly, that it impairs entrenched and carefully nurtured interests of people such as the insolvency practitioners, who have invested much time, effort, skill and money into the development of their commercial and concomitant interests; and thirdly, that it engenders social divisions that ultimately foster a culture of envious complaint which, left unchecked, can culminate in genocidal tendencies and civil war. In contrast to the multiracial impulse, which celebrates the use of race as a determinant, the values I have identified can only deplore the racial criterion as an evil, a necessary one, that they are forced to endorse.

Of course such a system does nothing to parade the supposed virtue of a system of social engineering designed to ‘create a black middle class’. Multiracialism, by contrast, serves this purpose admirably. However, this is precisely why multiracialism is racist and precisely why we should denounce it as unconstitutional. In our constitutional democracy, there is no place for race-cum-gender grids of the sort with which our society has become familiar. They constitute quotas pure and simple and—

A racial quota derogates from the human dignity and individuality of all to whom it is applied; it is invidious in principle as well as in practice. Moreover it can easily be turned against those it purports to help. The history of the racial quota is a history of subjugation, not beneficence. Its evil lies not in its name, but in its effects; a quota is a divider of society, a creator of castes, and it is all the worse for its racial base, especially in a society desperately striving for an equality that will make race irrelevant.

Why the Court has accepted multiracialism so enthusiastically is a complex question whose answer travels way beyond the scope of this paper. Here it is enough to observe that, by espousing multiracialism, mechanistic and inflexible as the system is, the Court has undermined the development of a genuine non-racial jurisprudence in our country. Under the latter philosophy, which is foreshadowed by aspects of our common law, recourse to race as a

---

116 James Myburgh depicts the horrific consequences that multiracialism can produce when taken to its logical conclusion. See ‘Race quotas: The Terrible Power of “Demographic Representivity”’, available at http://www.politicsweb.co.za/news-and-analysis/race-quotas-the-terrible-power-of-demographic-repr. He provides an excerpt from a Nazi party bulletin that is especially chilling: ‘There are more than 8,000,000 unemployed in Germany. Among them hundreds of thousands of Intellectuals. Nevertheless the German People have admitted hundreds of thousands of Jewish Intellectuals to the liberal professions. The same academic Jewish circles today thank Germany by lowering her in the estimation of the world. To check this, a demand will now be voiced to admit Jews to universities and to the professions of attorney and physician only in proportion to their numerical strength among the population of Germany.’ Later he quotes Goebbels as saying that ‘the Jewish question in Germany is not treated as a racial question [but] merely intended to return the Jewish element to that participation in public life and other activities which is in proportion to its quota within the general population.’ In a 1937 article in Die Transvaler, referenced by Myburgh, Hendrik Verwoerd, then editor, pursued the theme assiduously. ‘Legislation must gradually but purposefully ensure that each section of the population [Jews, Afrikaners and English] should, as far as practicable, enjoy a share of each of the major occupations, according to its proportion of the white population.’

117 See Charles Simkins ‘Genocide’, available at https://hsf.org.za/publications/hsf-briefs/genocide. Genocide Watch places South Africa at Level 6 on the Stanton Scale which ranges from one (mildest) to ten (worst). At this level a country is considered to be in a state of ‘Polarisation, with extremists driving groups apart, hate groups broadcasting polarising propaganda, and, possibly, laws forbidding intermarriage or social interaction.’


119 Inadvertence is not an explanation, it must be said. In each of the three cases in point, the Court was told that multiracialism was at the heart of the grid system.
criterion for decision-making would be permissible only if legitimate aims could reasonably be attained by no other means. Laurie Ackerman, a judge elevated to the Court at its inception, has written eloquently in support of such a system.\textsuperscript{120}

\textit{[N]o country where serious and systemic discrimination has occurred can – for administrative and other financial reasons – be expected to devise and administer a remedial system that has, in all cases, to consider the awarding of a remedy on a case-by-case, individual-by-individual basis. The logistics and costs might make this impossible. No legislation or comprehensive administrative measure can always operate with the surgeon’s scalpel, given the vast number of cases involved. It is, in appropriate circumstances, compelled to use a broader sword, but where an individualised remedy is reasonably possible it should be employed. … The overarching goal is not limited to establishing, progressively, a society in which the consequences of past discrimination are eliminated, but a society in which the dignity of all is equally respected.}

Since poverty and race are highly correlated, the integration of black people into society might as readily be achieved using Ackermann’s approach, and the disadvantages of an overt race-based structure – its capacity to create rancour among whites and triumphalism among blacks – will in the process be, if not eliminated, then certainly reduced. The fact that race-based remedial action is permissible under our Equality Clause does not mean that it must be employed as a tool of convenience. It should be utilised, as Judge Ackerman says, only when an individualised approach is impracticable, allowing always for the fact that, where appropriate, race-based information and statistics can be used as only one metric, among others, by which to evaluate the scheme.\textsuperscript{121} Such an approach would do justice to the notion that, while racial quotas are taboo, equivalently framed targets are not.

To some, the propositions I make and the distinctions I draw might seem to be trivial. Obviously I do not think so. I share the view that ‘splitting society into so many more or less fixed groups, all ceaselessly seeking privileges at the expense of all the rest, [makes] equality both meaningless and impossible.’\textsuperscript{122} A nation state cannot be built on such foundations and must, history teaches us, collapse into internecine conflict if the attempt is made.\textsuperscript{123} Our Constitution was devised to prevent, by such means as the law can command, so unhappy a result. The Court was charged with holding the fort but, traducing its trust, has meekly surrendered it to the structural and societal forces that are driven by race-consciousness.

The Court was wrong to be so acquiescent. As Sachs J said,\textsuperscript{124} ‘the state may not impose orthodoxies of belief systems on the whole of society’ and this is as true of racialism as of homosexuality, the topic with which he was concerned. ‘At the heart of equality jurisprudence,’ he rightly emphasised in the same case, ‘is the rescuing of people from a caste-like status and putting an end to their being treated as lesser human beings because they belong to

\begin{itemize}
  \item \textsuperscript{120} L Ackermann \textit{Human Dignity: Lodestar for Equality in South Africa} (2012) 358, emphasis supplied.
  \item \textsuperscript{121} Speaking of the United States, Bowen is quoted with approval in \textit{Bakke} for the following proposition: ‘While race is not in and of itself a consideration in determining basic qualifications, and while there are obviously significant differences in background and experiences among applicants of every race, in some situations race can be helpful information in enabling the admission officer to understand more fully what a particular candidate has accomplished – and against what odds.’ \textit{University of California Regents v Bakke} 438 US 265 (1978) 317 n51
  \item \textsuperscript{122} Martin van Creveld \textit{Equality: The Impossible Quest} (2015) 299.
  \item \textsuperscript{123} Surely no one needs a recital of such cases. What we do need, in our racially charged society, is to think much harder about them, and the genocide in Rwanda provides as good a spur as any for the process: see P Gourevitch \textit{We wish to Inform You that Tomorrow We will be Killed with our Families} (1998), a truly horrific tale.
  \item \textsuperscript{124} Cf National Coalition for Gay and Lesbian Equality & Another v Minister of Justice & Others [1998] ZACC 15, 1999 (1) SA 6, 1998 (12) BCLR 1517 (CC) para 137.
\end{itemize}
a particular group.'¹²⁵ Nor is there any actual need for the Court to be complacent. After a comprehensive survey of the role of race in society, Kenan Malik, perhaps the foremost authority on multiculturalism and identity politics, says ‘I am not suggesting that there may not be an evolved tendency to form in-groups and out-groups. But group conflict is not a given. The nature and strength of such conflict is historically contingent.’¹²⁶ It is as well to duck it if we can, but we will certainly not do so if we continue to divide people into groups framed by the immutable fact of skin colour, and dispose of them accordingly.

¹²⁵ Ibid at para 129.
The Silent Right: Environmental Rights in the Constitutional Court of South Africa

RUTH KRÜGER

ABSTRACT: Environmental rights are recognised in s 24 of the Constitution, but the Constitutional Court has not engaged with these in a meaningful manner in the last decade. Fuel Retailers, handed down in 2007, was the last case to engage s 24 fully. Although it used the concept of sustainable development to give content to the right, the case has been criticised for its economic focus and the lack of certainty in its approach. The resulting precedent may have been difficult to follow, creating a barrier for potential environmental cases in more recent years. However, a challenge prior to the interpretation of environmental rights in the court room is that environmental issues do not always make it to court. The three cases analysed in this piece reflect matters in which environmental rights were relevant, but neither argued by the parties nor raised by the Court. This may be attributed to the complexity of environmental matters, or to difficulties related to competing priorities and legal standing. More particularly, the challenges faced by the environment in getting to court are akin to the difficulties faced by children in having their rights vindicated – both must be represented by others. The three cases discussed in this article suggest that the Court has something of an environmental blindspot. This can be likened to the gender blindspots that have existed in past cases, and arguably still exist. Having made the case for an environmental blindspot, this article considers an array of tools that might address the problem. Internationally, independent environmental agencies and specialised environmental courts have been created, but these may not be the best fit for the South African context. Instead, this article suggests that existing legal tools may be used for the realisation of s 24’s environmental rights. The Constitutional Court may make use of its broad powers to pick up on environmental issues even where these have not been explicitly identified by the parties. While this may be contrary to established procedure, the Constitutional Court does adopt a flexible approach when it believes that specific rights or provisions in the Constitution are essential to ventilate and to resolve the issues raised in a matter. This flexibility may allow the court to bypass the challenges facing s 24, enabling the development of this important constitutional right. Finally, this article suggests a proactive approach to s 24, formalising an upper guardian for the environment to accelerate the development of the right’s content.

KEYWORDS: Section 24, environmental rights, sustainable development, complexity, legal standing, procedural rules, upper guardian

AUTHOR: BA(Hons) environmental science (Rhodes), MSc sustainability science (Lund); School of Law, University of the Witwatersrand. Email: ruth.m.kruger@gmail.com
ACKNOWLEDGEMENTS: Many thanks are due to Professor Stuart Woolman, who provided this piece with the perfect title. In addition, I am very grateful to Professor Salim Nakhjavani, Dr Emile Zitzke and Professor Woolman for extensive feedback on drafts of this piece.
I INTRODUCTION

24. Environment
Everyone has the right
(a) to an environment that is not harmful to their health or well-being;
(b) to have the environment protected, for the benefit of present and future generations, through reasonable legislative and other measures that
(i) prevent pollution and ecological degradation;
(ii) promote conservation; and
(iii) secure ecologically sustainable development and use of natural resources while promoting justifiable economic and social development.

It has been more than a decade since the Constitutional Court last engaged with environmental rights in a meaningful manner. In 2007, Fuel Retailers attempted to flesh out the content of s 24 using the concept of sustainable development. However, the absence of engagement in the interim should not be understood as an absence of Constitutional Court cases in which environmental rights were relevant. In fact, s 24 could have been applied in a range of cases. Taking creative liberty with the influential work of Rachel Carson, it appears that s 24 has become a silent right.

The ongoing silence exists despite South Africa’s clear and commendable commitment to the global environmental consciousness that was created by actors such as Carson. In fact, following the recognition of sustainable development at global moments such as the Brundtland Commission in 1987, and the UN Conference on Sustainable Development in 1992, South Africa was the host of the World Summit on Sustainable Development in 2002. Furthermore, South Africa is a signatory to a number of international environmental treaties, and is constitutionally obliged to consider both binding and non-binding international law in any case.

The silence surrounding s 24 is therefore a strange phenomenon, and all the more so given the number of cases since Fuel Retailers in which environmental rights have been implicated. Nevertheless, there has been no significant further engagement with s 24 – at most these
cases contained brief references to the right. For example, the case of Mazibuko turned on the right to water in terms of s 27,9 but the close link of this right to environmental rights was not acknowledged. Similarly, Merafong Municipality concerned the pricing of water,10 but the relevance of this issue to s 24 was not discussed. Water was also at the centre of Aquarius Platinum,11 and though this case did mention s 24, it did not take the matter further. Several cases have discussed mining without considering its environmental impact, such as Bengwenyama,12 and AgriSA.13 Maccsand14 did slightly better in its discussion of mining, as it mentioned both s 24 and the National Environmental Management Act 107 of 1998 (NEMA). However, Maccsand fell short of providing substantive guidance on the application of environmental rights in the context of mining. Then, the Biowatch case – despite being ‘only about costs awards’15 – clearly acknowledged that these costs awards are connected to ‘environmental justice’.16 But it did no more than place the full text of s 24 in a footnote, without discussing its relevance.17 Various cases have considered the powers of different levels of government with regard to town planning and land use planning, without acknowledging the implication of such decisions for land management and therefore environmental rights.18 Recently, the Constitutional Court in NSPCA19 acknowledged that a link exists between animal welfare and s 24, but did not explain the nature or importance of the link.

One possible source of the silence is that Fuel Retailers – as the most significant engagement with s 24 to date – may have failed to provide certainty as to the content of the right. While the Court did engage meaningfully with sustainable development by suggesting that this concept be used to give meaning to s 24,20 its approach is vague and confusing. As further discussed below, the precedent created is arguably difficult to follow in further cases.

---

10 Merafong City Local Municipality v AngloGold Ashanti Ltd [2016] ZACC 35, 2017 (2) SA 211 (CC)(‘Merafong Municipality’).
11 Minister for Environmental Affairs & Another v Aquarius Platinum (SA) (Pty) Ltd [2016] ZACC 4, 2016 (5) BCLR 673 (CC)(‘Aquarius Platinum’).
12 Bengwenyama Minerals (Pty) Ltd & Others v Genorah Resources (Pty) Ltd & Others [2010] ZACC 26, 2011 (4) SA 113 (CC)(‘Bengwenyama’).
13 Agri South Africa v Minister for Minerals and Energy [2013] ZACC 9, 2013 (4) SA 1 (CC)(‘AgriSA’).
14 Maccsand (Pty) Ltd v City of Cape Town & Others [2012] ZACC 7, 2012 (4) SA 181 (CC)(‘Maccsand’).
16 Ibid at para 9.
17 Ibid at fn 9.
20 Fuel Retailers (note 1 above) paras 44–62.
To help fill the gap, various commentators have suggested alternative or complementary approaches that might give greater substance and coherence to s 24.21 Tladi and Feris each devised a flexible set of tools which require the conscious choice of a normative direction for interpreting s 24.22 De Wet and Du Plessis have suggested that we might find useful principles for the application of environmental rights in international legal frameworks.23 Zooming in rather than out, Fuo developed an argument for a greater role for local government.24 Du Plessis cautioned that environmental rights cannot be understood, or responsibly applied, without an understanding of poverty.25 Recently, Blackmore connected the public trust doctrine with environmental rights interpretation.26

These suggestions by commenters are likely to prove invaluable when the Court is called upon to give meaning to s 24. However, prior to any court room engagement with environmental rights, a larger issue is that environmental issues may not be making it to the courts. In the cases mentioned above, environmental rights were absent from the pleadings. Further, the Court failed to identify the relevance of s 24, despite its wide powers and the fact that at times the Court has provided relief beyond what has been pleaded before it. Much like gender blindspots in other cases, the Court may have an environmental blindspot. The cases mentioned above might have been argued differently if environmental rights had been pleaded and considered. This could have changed their outcome, or else would have contributed to a stronger jurisprudence to protect people, nature and sustainability in the future.

In developing the argument in this paper, I will give a brief synopsis of Fuel Retailers and where I see possible failings. I will then reflect upon the problem I see today – the post-Fuel Retailers silence surrounding s 24 – by taking a closer look at three more recent cases in which s 24 would have been relevant. I then make suggestions about the effect that a consideration of s 24 might have had on these three matters. I will argue, however, that any failings of Fuel Retailers are only part of the story. A greater problem is that environmental rights cases simply do not appear to be reaching the Court. I will briefly consider possible explanations for this absence on the docket. More space will be devoted, however, to suggestions as to how institutions could be created or adapted to expand the protection offered to South Africans through environmental rights. Foreign legal systems provide some assistance, but I ultimately suggest alternative approaches that may be better suited to South Africa’s particular context.


22 Feris 1 (note 21 above); Tladi 1 (note 21 above).

23 de Wet & du Plessis (note 8 above).


25 Du Plessis (note 21 above).

II THE CONSTITUTIONAL COURT’S LAST MEANINGFUL ENGAGEMENT WITH ENVIRONMENTAL RIGHTS

As already noted, the Court last engaged with environmental rights in meaningful manner in Fuel Retailers. In brief, this case concerned an application which was granted by the provincial environmental authorities in Mpumalanga for the opening of a filling station in White River. The decision was challenged by the Fuel Retailers’ Association of Southern Africa on the grounds that the environmental authorities did not consider the socio-economic impact of the proposed filling station.

The matter, which was decided in 2002, turned on an Environment Impact Assessment (EIA) which considered the feasibility of the proposed filling station, as well as various social issues: noise, visual impact, traffic, the effect on municipal services, safety and crime, and cultural and historical sites. The papers before the Court included a geotechnical report which concluded that a subterranean aquifer beneath the proposed filling station might need protection from pollution if the Department of Water Affairs and Forestry thought so. In that case an impermeable base layer should be laid down beneath the station. Crucially for this case, however, the EIA did not consider the potential economic impact on existing filling stations in the area.

The applicants’ case before the Court was built on this failure to consider the economic impact the filling station would have, primarily, on other filling stations in the area. What was supposedly lacking was summarised as the consideration of need, desirability and sustainability. The environmental authorities countered that these issues were part of the analysis by the town planning authority for rezoning eight years previously. And so, battle lines were drawn.

Writing on behalf of the majority, Ngcobo J started by providing a useful synopsis of the development of the concept of sustainable development. In turning to the scope of s 24, he identified both the inherent tension between development and environmental protection, and the need to reconcile the two for the purposes of sustainability. However, a better explanation of both sustainable development and s 24 was needed. Sustainable development is inherently complex, made up as it is of three interlinked pillars and overlain with systems dynamics. There is no one right pathway to achieve sustainability, and to insert the concept as an objective measure has little meaning. In fact, Tladi argues, it leaves the

27 Fuel Retailers (note 1 above) para 1.
28 Ibid at para 5.
29 Ibid at paras 9, 14.
31 Ibid.
32 Ibid at paras 15–17, 22–24.
33 Ibid at para 15–17.
34 Ibid 22–24.
35 Feris 1 (note 21 above) at 240–244. My account draws on Feris’ summary.
36 Fuel Retailers (note 1 above) at paras 44–62.
37 Feris 1 (note 21 above) at 250.
Court open to value-based decision-making – without clear or conscious acknowledgement of these potentially unhelpful values.

Nevertheless, Ngcobo J went on to apply his understanding of sustainable development to the facts. He recognised the relationship between socio-economic factors and the environment, and cautioned that a proliferation of filling stations in one area could have a negative impact on the environment.\(^40\) In particular, he found that if there were too many filling stations some would close down, and the land would need to be rehabilitated.\(^41\) In terms of administrative structures, he found that the assessment that was done by town planning authorities had been insufficient. This was both because it had been done over eight years prior to the creation of the filling station,\(^42\) and because those authorities fulfilled a different function than that required of environmental authorities.\(^43\) Ultimately, he found for the Fuel Retailers Association and directed the matter to be sent back to the High Court for the relevant socio-economic issues to be properly considered.\(^44\)

Ngcobo J’s analysis did not engage with sustainability in an appropriate manner, which suggests that his view of the related concept of environmental rights is limited. His primary argument revolved around the economic sustainability of other filling stations in the area.\(^45\) While relevant, it cannot be the only consideration in a sustainability analysis, as it is only one element of the complex human-environment system being considered for land-use change.\(^46\) Further, while Ngcobo J’s concern about rehabilitation may appear to refer to the natural world, I would argue that it also turns on economic considerations. Otherwise the argument is nonsensical. The environmental damage leading to a need for rehabilitation does not spring into being when a filling station closes. It is present throughout the life of the filling station, but generally it is only dealt with when the land needs to be prepared for another use. However, it is true that there are cost implications associated with land rehabilitation which only come up when a filling station closes – that is surely what Ngcobo J was referring to. His concern is economic, not ecological.

In the minority judgment, Sachs J’s analysis was more holistic. It made explicit the need to consider social and environmental impacts alongside economic impacts.\(^47\) However, this is only when there is in fact a negative impact on social and environmental systems, and Sachs J found that none existed.\(^48\) This weakness in his judgment reflects his limited understanding of social and environmental effects. It considers only the immediate, local level impact of a filling station on the land, water and air. This analytical framework falls short of the systems understanding needed for sustainability.\(^49\) Blackmore suggests that while mining rights are not public goods

\(^{40}\) Fuel Retailers (note 1 above) paras 71–83.
\(^{41}\) Ibid at para 71.
\(^{42}\) Ibid at paras 94–96.
\(^{43}\) Ibid at paras 84–85.
\(^{44}\) Ibid at paras 105–106.
\(^{45}\) Tladi 1 (note 21 above) at 258.
\(^{47}\) Fuel Retailers (note 1 above) at para 113.
\(^{48}\) Fuel Retailers (note 1 above) at 114.
that would be subject to the public trust doctrine, they must be seen holistically as there are associated impacts of mining that do impact other public goods. Similarly, filling station themselves may have a low impact on social and environmental systems. However, when they are viewed holistically they enable fossil fuel use, which has been clearly linked to climate change and general air quality problems. In South Africa it is often poor communities that experience the health risks of mines and factories producing the fuel used at filling stations like the one in question. Sachs J had the chance to make a point about environmental rights in the context of fuel production, and more generally about South Africa’s development trajectory, but he did not do so. He did acknowledge the irony of an environmental issue being raised by such a polluting industry, but failed to tie this insight into his more general analysis.

Ultimately, the guidance provided by Fuel Retailers regarding the application of s 24 is limited. As a reasoning tool, sustainable development is useful but vague, and raises more questions than it answers. In this case it appears to have allowed for an understanding of sustainability that does not depart from the economic focus of the status quo. Tladi and Feris reached this conclusion over a decade ago. Today, however, we can see the effect of Fuel Retailers on subsequent cases – or the lack thereof. In the next section, I will briefly unpack three cases in which s 24 was not considered, although it was relevant. I will then offer some thoughts as to how the outcome might have been different if s 24 had in fact been considered.

III AFTER FUEL RETAILERS: THREE CASES WITH ENVIRONMENTAL RELEVANCE

A Lagoonbay

Lagoonbay dealt with an application for the subdivision and rezoning of a large piece of land for a property development in the Southern Cape. The application was considered together with ss 16 and 25 of the Land Use Planning Ordinance 15 of 1985 (LUPO), which gave provinces the authority to refuse subdivision and rezoning decisions for land, respectively. Following LUPO, applications by Lagoonbay Lifestyle Estate (Pty) Ltd (Lagoonbay) for subdivision and rezoning were granted by the local authority but then refused by the provincial authority. Lagoonbay challenged the decision of the provincial authority, claiming it was acting ultra vires in terms of the functional competencies of provinces and municipalities set out in the Constitution.

The application was dismissed in the High Court but upheld in the Supreme Court of Appeal. In the Constitutional Court, the application was upheld in part and dismissed in part. In a unanimous judgment written by Cameron J, the Court declined to consider the

50 Blackmore (note 26 above) at 196–197.
53 Fuel Retailers (note 1 above) at para 109.
54 Tladi 1 (note 21 above).
55 Feris 1 (note 21 above).
56 Lagoonbay (note 18 above)
57 Ibid at para 1.
58 Constitution, Schedules 4–5.
ENVIRONMENTAL RIGHTS IN THE CONSTITUTIONAL COURT OF SOUTH AFRICA

It found a direction in the 1988 Circular which transferred powers from municipalities to provinces, and rezoning was among these. The provincial authority was therefore vindicated as far as rezoning was concerned. However, the same was not true for their subdivision decision. The province itself issued scheme regulations in terms of LUPO in 1988, and these left no room for provincial control of subdivision applications. There was an amendment of the scheme regulations in 2009, but this gave municipalities the power to decide whether they would make a subdivision decision, or whether it would be left to the province. Accordingly, the Court found that the provincial authority had not had the authority to refuse Lagoonbay’s subdivision application, but had had the authority to refuse the rezoning application.

B AgriSA

AgriSA dealt with the question of whether the Mineral and Petroleum Resources Development Act 28 of 2002 (MPRDA) allowed for the expropriation of mineral rights that had been held under the Minerals Act 50 of 1991. Before the MPRDA, mineral rights belonged to the owners of a piece of land, although they were severable. In terms of the MPRDA, however, mineral rights vested in the state, and there were transitional arrangements that allowed previous right holders (now holders of ‘old order rights’) to convert their rights so that they were recognised under the MPRDA. Seeking to defend the mineral rights of its members, Agri South Africa (Agri SA) launched a court application, using the company Sebenza (Pty) Ltd (Sebenza) as a test case. Sebenza had bought coal rights from the liquidators of Kwa-Zulu Collieries, but could not convert their old order rights as the colliery could not pay the fees to apply for prospecting and mining authorisations under the MPRDA. Agri SA asserted that the MPRDA allowed for the expropriation of mineral rights.

The application was successful in the High Court, but failed in the Supreme Court of Appeal. In the Constitutional Court, the majority judgment written by Mogoeng CJ found that, while there was a deprivation of rights, it did not amount to expropriation. The Court cautioned that s 25 property rights should not be overemphasised in view of the need for transformation. It also listed the requirements for expropriation: (i) compulsory acquisition of the rights in property by the state, (ii) for a public purpose or in the public interest, and (iii) subject to compensation. The analysis failed at the first requirement, as the state did not acquire the mineral rights according to Mogoeng CJ. It became the holder, but only to enable their equitable distribution and use. In the specific case of Sebenza, the Court held

59 Lagoonbay (note 18 above) at para 40.
60 Ibid at paras 50–52.
61 Ibid at para 53.
62 Ibid at para 54.
63 AgriSA (note 13 above).
64 AgriSA (note 13 above) at paras 7–12.
65 Ibid at paras 2, 27–30.
66 Ibid at paras 16.
67 Ibid at paras 13–15.
68 Ibid at paras 18–20.
69 Ibid at paras 62–64, 70.
70 Ibid at para 67.
71 Ibid at paras 68–71.
that the transitional arrangements would have allowed it to convert its old order rights – only its own financial position prevented this conversion.\textsuperscript{72} However the Court did consider the importance of the various purposes of the Act: to facilitate equitable access to the mining industry, to promote sustainable development, and to eradicate discriminatory practices in the mining sector.\textsuperscript{73}

C Merafong Municipality

\textit{Merafong Municipality}\textsuperscript{74} came to Court following a decision by Merafong City Local Municipality (Merafong) to add a surcharge to the industrial and domestic water use of AngloGold Ashanti Limited.\textsuperscript{75} This action by the municipality was in accordance with their exclusive constitutional competence in terms of s 156(1) of the Constitution, read with s 1 of the Water Services Act 108 of 1997.\textsuperscript{76} However, AngloGold appealed to the Minister of Water Affairs and Forestry (the Minister), who was empowered to decide such an appeal in terms of s 8(4) of the Water Services Act.\textsuperscript{77} The Minister overturned the surcharge levied for industrial use, and ruled that the domestic water use tariff should be negotiated by the relevant actors.\textsuperscript{78} Merafong subsequently received legal advice, however, to suggest that they did not have to comply with the Minister’s ruling.\textsuperscript{79} They therefore threatened to turn off AngloGold’s water unless they paid the surcharge, and the company had no choice but to comply.\textsuperscript{80} They did, however, launch a court application to compel Merafong to comply with the Minister’s ruling.\textsuperscript{81}

The application was upheld in both the High Court and the Supreme Court of Appeal.\textsuperscript{82} Merafong gained leave to appeal to the Constitutional Court. Here, Merafong contended that where a public official makes a decision that is outside the scope of their powers, this decision may be ignored until there is an attempt to enforce it, and at this point the nullity of the decision may be raised.\textsuperscript{83} Cameron J, writing for the majority, disputed the Supreme Court of Appeal’s finding that a collateral challenge may only be raised by an individual, not an organ of state, and found that reactive challenges have always been treated flexibly in our law, depending on context.\textsuperscript{84} Following the lower courts in considering the cases of \textit{Oudekraal}\textsuperscript{85} and \textit{Kirland},\textsuperscript{86} Cameron J then went on to confirm that Merafong should have either followed the Minister’s decision or challenged it in court, rather than resorting to self-help – this conclusion flows from the Court’s understanding of good constitutional citizenship.\textsuperscript{87} That said, Cameron J

\begin{footnotesize}
\begin{enumerate}
\item Ibid at para 72.
\item Ibid at paras 64–65, 73.
\item \textit{Merafong Municipality} (note 10 above).
\item Ibid at para 5.
\item Ibid at paras 1 and 2.
\item Ibid at paras 2 and 9.
\item Ibid at para 1.
\item Ibid at para 10.
\item Ibid.
\item Ibid at para 13.
\item Ibid at paras 14–15.
\item Ibid at para 17.
\item Ibid at paras 55–56.
\item \textit{Oudekraal Estates (PTY) Ltd v City of Cape Town} [2004] ZASCA 48, 2004 (6) SA 222 (SCA) (‘\textit{Oudekraal}’).
\item \textit{MEC for Health, Eastern Cape v Kirland Investments (PTY) Ltd} [2014] ZACC 6, 2014 (3) SA 481 (CC) (‘\textit{Kirland}’).
\item \textit{Merafong Municipality} (note 10 above) at paras 59–64.
\end{enumerate}
\end{footnotesize}
went on to consider whether Merafong’s belated ‘conditional counter-application’ did raise a substantive challenge to the Minister’s decision.\textsuperscript{88} Ultimately, he held that Merafong should have the chance to justify its delay in seeking judicial recourse and so remitted the matter to be considered again by the High Court.\textsuperscript{89}

**IV \hspace{1em} WOULD SECTION 24 HAVE CHANGED THE OUTCOME OF THESE CASES?**

The absence of judicial appraisal of environmental rights in the three cases above is not important, however, unless it leads to some harm. In this section, I will therefore assess whether the consideration of s 24 would have changed the outcome of the cases. First, I will set up an argument regarding the relevance of s 24 to the three cases described above – all of which occurred after *Fuel Retailers*. I will then make suggestions as to how environmental rights might have been applied. Finally, even where s 24 would not have altered the outcome of a case in a meaningful manner, I will suggest that application of s 24 would have assisted in building a stronger environmental jurisprudence in Constitutional Court litigation.

In *Lagoonbay*, the issue was one of a large land development, and land use is an environmental issue. Inappropriate land use may lead to land degradation, an issue of such importance that one of the UN’s three major environmental conventions is dedicated to it.\textsuperscript{90} In the proceedings before court here, it appears that the environmental link was acknowledged to some extent, as a local environmental organisation was admitted as respondent (the Cape Windlass Environmental Action Group). However, the organisation’s submissions do not appear to have had a meaningful impact on the Court’s decision as none of these interventions were mentioned in the judgment. Further, one of the arguments put forward in favour of the Provincial Minister was that the size of the development meant it needed to be considered in terms of strategic provincial planning. This argument is in line with the approach favoured by the Department of Environmental Affairs (DEA). The DEA has in the past called for larger and land-focused Strategic Environmental Assessments when planning across a larger area, as opposed to smaller and project-focused Environmental Impact Assessments.\textsuperscript{91} However, the Provincial Minister’s argument was not taken up by the Court, or placed in the context of s 24. If it had been, then the Court might have considered environmental factors to arrive at its decision. Rather than considering government competences as they are, the Court might have considered what they should be. It might have considered what capacities are necessary for responsible decisions to be made about land management, and what level of government has those capacities. This approach might have favoured provincial competences over subdivision and rezoning where a large piece of land is concerned – in line with DEA policy, as described above. Alternatively, there would at least have been a more extensive enquiry into what is needed for environmentally responsible decisions, perhaps considering issues of scale, resources and relevant skills.

*AgriSA* concerned mining rights, an area of economic activity whose potential environmental impacts are large – both in terms of ecological and human systems.\textsuperscript{92} Some of the social

---

\textsuperscript{88} Ibid at para 65.

\textsuperscript{89} Ibid at paras 66–82.

\textsuperscript{90} United Nations ‘Convention to Combat Desertification’ (n.d.), available at https://www.unccd.int/.


\textsuperscript{92} T Madihlaba ‘The Fox in the Henhouse: The Environmental Impact of Mining on Communities’ in D McDonald (ed) *Environmental justice in South Africa* (Athens, Ohio: Ohio University Press, 2002)
challenges led to the need for the MPRDA, which aims to reform the mining industry to ensure that all South Africans benefit from it.\textsuperscript{93} The question of who has access to mining rights was certainly relevant in this case – as emphasised by the Court – but so is the question of how the mining is done. If the exercise of mining rights means that nearby communities are constantly subjected to dust or polluted air that leads to illness such as asthma,\textsuperscript{94} or if acid mine drainage pollutes groundwater,\textsuperscript{95} or if migrant labour leaves hollow, unsustainable settlements behind in rural areas,\textsuperscript{96} then there is a clear infringement of environmental rights. These problems were not at issue before the Court, but what was at issue was the constitutionality of the state’s role in relation to mining rights. The matter specifically engaged the process whereby people gain access to mining rights under the new legislation. I would suggest that a proactive Court would have considered this constitutionality in terms not only of property rights, but also in terms of environmental rights. Had the Court taken this approach, it may nevertheless have dismissed the necessity of compensation, but might have suggested more meaningful environmental rights requirements in the authorisation processes for mining rights. Had the state been asked to take on a stronger environmental rights role, the justification of its powers under the MPRDA would have been stronger – particularly given the clear reference to environmental protection in the preamble of the Act. If the state has control over mineral rights to reform the mining sector, it need not stop at patterns of ownership and employment. It is similarly in the interests of justice for mining not to infringe on environmental rights, and the Court might have recognised the potential of the MPRDA as a tool for this. This question may not have been directly before the court, but I would suggest that it is relevant to the constitutionality of the transitionary arrangements in the MPRDA. A proactive court might have incorporated the issue into the judgment, or at least provided an \textit{obiter dictum} that would have strengthened an environmental rights jurisprudence that is currently very thin.

\textit{Merafong Municipality} involved a surcharge on water. Water is an important environmental good, particularly in South Africa, which receives just over half of the world’s average rainfall,\textsuperscript{97} at 497 mm per year.\textsuperscript{98} These challenges have recently been thrown into stark relief in Cape Town, which may yet become the first major city in the world to run out of water.\textsuperscript{99} Therefore, the management of water in South Africa is both important and quite obviously connected to environmental rights. After all, a lack of sufficient water is clearly harmful to one’s health and well-being, and the nature of its use may lead to pollution and ecological degradation. \textit{Merafong Municipality} deals with water directly, but appears to have a blindspot regarding the implications of its decision for environmental rights. The industrial surcharge was summarily

\begin{itemize}
  \item AgriSA (note 13 above) paras 25–26.
  \item W Beinart ‘A Century of Migrancy from Mpondoland’ (2014) 73(3) \textit{African Studies} 387.
  \item Rainfall is measured using rainfall gauges at a fixed location over a period of time. Average annual precipitation is calculated as the mean of rainfall depth from all available rainfall data worldwide in a given year. It is a very broad measure, but here may usefully demonstrate South Africa’s low rainfall levels.
\end{itemize}
dismissed, while the levy for domestic water use was left open for negotiation. This is an anomaly because of the well-established causal link between industrial water use and pollution, and therefore environmental rights. A tax or surcharge on industrial water use might be used for better environmental management, or it might be a so-called ‘sin tax’ designed to improve the efficiency and sustainability of industrial water use. But neither the Minister nor the Court considered these possibilities. On the other hand, the taxation of domestic water use was allowed, despite its relationship to the right to water in s 27 of the Constitution. It would have been useful and appropriate for the Court to place this decision in the context of s 24. A rights-based consideration might have led the court to alter its decision so as to allow a surcharge on industrial water but not domestic water. Failing this, the question of environmental impact would have been more clearly established as a relevant factor where decisions on water are concerned, and particularly where they relate to mining.

In summary then, a consideration of s 24 might have changed the ratio decidendi of the three cases that are unpacked here. It might have introduced new focus points. It might even have changed the outcome. The Lagoonbay Court might have found both subdivision and rezoning of land to be provincial competences, or would at least have considered the strategic competences necessary to make broader decisions about land use. In AgriSA, environmental rights would have most likely only strengthened the finding that the MPRDA does not allow for expropriation, adding to the need for government oversight of mineral rights. A conscientious Court might have questioned the requirements needed for the granting of a mining authorisation, and ordered that these be reconsidered in terms of s 24 – or at least remarked upon this need in an obiter dictum. This might have improved the sustainability of mining in South Africa, or at least would have strengthened the environmental rights jurisprudence, adding to the protection of those whose rights will be violated in the future. In Merafong Municipality, the Court would likely still have remitted the case to the High Court. But it might have done so with the proviso that water management – and particularly industrial water use – should be considered in terms of s 24. It might also have established environmental impact as a factor to consider in future cases surrounding the management of water.

V WHY THE SILENCE OF SECTION 24?

So far, I have sought to demonstrate that proper consideration of environmental rights is missing from the Court’s deliberation in cases that have a bearing on environmental issues such as land use, water supply and the effects of mining. I have suggested that the absence of a consideration of s 24 is a function of Fuel Retailers’ vague construction of the right, which allows for a narrow focus on economic sustainability. However, a more significant problem may be that environmental cases are not making it to the Court. In this section, I will briefly consider some of the possible reasons for this silence: complexity, competing priorities and standing.

Regarding the complexity of environmental rights, the concept of sustainability is novel and continues to evolve. It was in the early 1980s that the concept of sustainable development

100 Merafong Municipality (note 10 above) at para 1.
103 Kates (note 39 above) at 19449.
emerged with the development of several international policy documents such as the Brundtland Report called ‘Our Common Future’.\textsuperscript{104} It took several more decades for a scientific field to be established. A mere eight years ago, in 2011, Bettencourt and Kaur’s study brought together evidence from the publication record to demonstrate that the practice of sustainability science did in fact exist.\textsuperscript{105} This was just before the three cases discussed above were decided, in 2013, 2014 and 2016, respectively. Bettencourt and Kaur’s study also pointed to the fast-growing nature of the practice – not yet a field – which was extending into non-traditional geographic and thematic areas in response to new environmental challenges.\textsuperscript{106} A court may not have full comprehension of the complexities of the practice, and would likely be hesitant to engage with environmental issues. Further, the legal framework surrounding sustainability – environmental rights – is nebulous,\textsuperscript{107} and novel not only in South Africa but across the world.\textsuperscript{108} It is new not only in jurisprudence, but in public discourse, and that may be part of the reason that neither the Court nor the parties to the three cases outlined above raise or address environmental rights arguments.

If there is still comparatively little recognition of or knowledge about environmental issues, this lack of awareness likely extends to the connection between ecological and social sustainability.\textsuperscript{109} Courts – and parties – may see the social side of an issue, but not the relevance of its environmental implications. Further, they might not fully appreciate the importance of environmental issues, if they view these in isolation and not as components of a complex socio-ecological system.\textsuperscript{110} A more system-oriented view of environmental issues would recognise that these issues must be resolved for the proper functioning of human systems,\textsuperscript{111} and therefore might give them greater weighting. However, such thinking has not been fully mainstreamed.\textsuperscript{112} It may be particularly challenging in a context such as South Africa, where poverty and inequality\textsuperscript{113} mean that there are many competing priorities.

In this regard, it may be useful to consider a parallel to gender mainstreaming. Courts have traditionally failed to recognise the gendered nuances in cases, and have therefore ruled incorrectly at times. One example is the standard of reasonableness used to test for a legal duty. It was traditionally dubbed the ‘reasonable man’ test, although it was supposedly objective. Today, the test refers to the ‘reasonable person’, but it has been argued that it has not lost its


\textsuperscript{105} Ibid.

\textsuperscript{106} Ibid.

\textsuperscript{107} Feris 2 (note 21 above) at 38–39.

\textsuperscript{108} de Wet & du Plessis (note 8 above)

\textsuperscript{109} The thinking surrounding the connection between social and ecological sustainability is novel and evolving. For a new approach see in J Buckles Education, Sustainability and the Ecological Social Imaginary: Connective Education and Global Change (Palgrove Macmillan, 2018).


\textsuperscript{111} Kates (note 39 above) at 19450.


\textsuperscript{113} In 2014, the World Bank measured South Africa’s GINI coefficient at 0.63. World Bank ‘GINI index (2019) (World Bank estimate)’, available at https://data.worldbank.org/indicator/SI.POV.GINI?locations=ZA.
gendered nature.\textsuperscript{114} To consider the test objectively, when in fact it is gendered, is arguably an ongoing failure of the courts in identifying all relevant nuances of a case. Another example of a gender blindspot is the case of Masiya.\textsuperscript{115} Here, the Court extended the definition of rape to include anal penetration of a female, but refused to make the definition gender neutral. Phelps and Kazee have argued that this decision sends the highly problematic message that females are both more vulnerable and more deserving of protection, and further that the ruling violates the equality provision of the Constitution.\textsuperscript{116} The Court here considered the effect of its ruling on the common law, but did not see the connection of this ruling to gender issues. The legislature has since intervened to extend the definition of rape,\textsuperscript{117} but the fact remains that this recent case demonstrates a clear gender blindspot at the Court. Similarly, I have suggested here that the Court has an environmental blindspot. Where there are environmental ramifications in the cases pleaded before it, they may not be identified by the Court, or indeed the parties. The resulting silence was seen in Lagoonbay, AgriSA and Merafong Municipality.

The final reason for the absence of s 24 from courtrooms is connected to legal standing. It has been suggested that the vindication of environmental rights is limited by the fact that the environment cannot go to court.\textsuperscript{118} It is necessary for humans to go to court on behalf of the environment, and this is reflected in the construction of environmental rights. Even in Ecuador, the first country to give the environment itself subjective legal rights,\textsuperscript{119} the vindication of these rights must be at the hands of those of us with opposable thumbs. There is simply no other way. South Africa, by contrast, has taken an anthropocentric approach to environmental rights – the environment is protected through its connection to humans.\textsuperscript{120} This is an open acknowledgement of the human role in defending environmental rights, and underlines the fact that the environment depends on human action for its protection. However, the fact remains that access to court for environmental issues may be hampered as they are filtered through human interests. Even where there is an attempt to take environmental issues to court, human actors may lack standing to do so. This additional difficulty will be unpacked below.

In considering this final point, it may be useful to draw a parallel to children’s rights. Children must be assisted in going to court, much as the environment must be. Their legal capacity is limited at best, while the environment is not a legal person in South African law, and therefore has no legal capacity at all. Both children and the environment may have legitimate interests in a case, but their interests are unlikely to be taken into account unless recognised legal persons intervene on their behalf. For both, this status may be disempowering. In South Africa, lawmakers appear to have been very aware of this challenge as regards children, specifying that a child’s interests are ‘of paramount importance in every matter concerning the child’, per s 28(2) of the Constitution. To meet the imperative, the Children’s Act sets up a


\textsuperscript{115} Masiya v Director of Public Prosecutions, Pretoria (The State) &Another [2007] ZACC 9, 2007 (5) SA 30 (CC) (‘Masiya’)


\textsuperscript{117} Criminal Law (Sexual Offences and Related Matters) Amendment Act 32 of 2007.

\textsuperscript{118} T Murombo ‘Strengthening locus standi in Public Interest Environmental Litigation: has Leadership Moved from the United States to South Africa?’ (2010) 6(2) Law, Environment and Development Journal 165, 172.


\textsuperscript{120} Feris 2 (note 21 above) at 49.
supporting role for social workers, a family advocate and psychologists,121 and further creates the institution of Children’s Courts.122 In addition, the High Court is established as the upper guardian of children’s rights in terms of s 45(4) of the Children’s Act. In this role, the High Court must raise issues that are in the interests of children whether or not they have been raised by the parties. Had the Court acted in such a manner in the three cases examined above, it might have identified the relevant environmental issues, even though the parties had not. As it is, however, the environment failed to gain access to court.

There are therefore a number of barriers that environmental issues may face in reaching the courts. Among these are the novelty and complexity of environmental issues, the fact that South African society faces many competing priorities, and the issue of legal standing. In attempting to understand the silence surrounding s 24 rights, I have suggested that useful parallels may be drawn to the treatment of gender in the courts, and the challenges of ensuring access to court for children.

VI CREATING LEGAL CONDITIONS TO BREAK THE SILENCE

In this section, I will consider tools that might be used to encourage greater access to court for environmental rights issues. I will first consider international approaches, and then make suggestions as to how some of the existing legal tools in South Africa might be useful.

A Environmental agencies and legal standing

The first international approach that deserves attention is the creation of a specified agency with the purpose of promoting appropriate environmental governance. One of the best-known examples of this approach is possibly the Environmental Protection Agency (‘EPA’) in the USA,123 and similar institutions exist in other countries, such as the Umweltbundesamt (‘UBA’) in Germany.124 Theoretically, the existence of such a body might be a solution to the legal standing challenge mentioned above, improving access to court for environmental issues. This is the case not because human mediators are no longer necessary, but rather because there is a specified group of human mediators who are mandated to ensure that environmental issues are considered.

However, in practice it appears that environmental agencies may not function directly to address the issue of legal standing. In the case of the EPA, this organisation appears to act primarily in the legislative space,125 and by achieving settlement agreements with groups challenging environmental interests.126 While this commendable work might be usefully emulated in other countries, it may not assist in defeating the challenge of legal standing.

This point finds further support in one of the USA’s most prominent cases on the environment and legal standing, Massachusetts v EPA.127 The case followed a decision by the

122 Ibid at Chapter 4.
123 Available at https://www.epa.gov/.
124 Available at https://www.umweltbundesamt.de/en.
126 Examples are the EPA’s agreements with Chevron regarding safety at its refineries, and Fiat Chrysler regarding their infringements of air quality regulations. See: EPA ‘Civil Cases and Settlements’ (2018), available at https://cfpub.epa.gov/enforcement/cases/.
127 549 U.S. 497 (‘Massachusetts’).
EPA that it was not the body with the authority to regulate greenhouse gases, and that it declined to do so in any case, due to the scientific uncertainty surrounding climate change. Various states, cities and private organisations challenged this decision, and the case reached the Supreme Court of the USA in 2007. This court found that the state of Massachusetts did have legal standing, in part because of the ‘special solicitude’ that must be granted to federal states in recognition of their quasi-sovereignty. This is a very important case regarding legal standing in environmental cases. It has been suggested that it loosens the formal standing test in the USA by emphasising purpose over form and lowering the threshold for sovereign states to achieve standing, particularly in the environmental context. Further, the case has been used in constructing an argument for the relaxation of standing rules so that the interests of future generations can be better protected. It is notable, however, that the EPA – the specified environmental agency – is the respondent, not the appellant. While Massachusetts allowed for a development of standing jurisprudence in the USA which may improve environmental governance significantly, this was not due to an intervention by the EPA. Rather, it was an action against the EPA with unexpected consequences.

It appears therefore that the existence of an environmental agency does not necessarily improve access to court for environmental issues. However, the principles surrounding legal standing in the USA are not identical to those in South Africa. It may not be advisable to treat lessons drawn from the experience in this country as entirely transferable. The way that standing functions in the USA is that three criteria must be met: ‘(1) a concrete and imminent injury (2) caused by the challenged action that would (3) likely be redressed by a favourable decision.’ This test was originally interpreted so that it was far easier to establish standing where the injury is personal in nature, and later remained a significant barrier to climate litigation because of difficulties in proving causation and redressability. Despite a proliferation of private environmental actions in the USA and the expectation of positive developments in this area, the standing test remains a barrier to individuals and interest groups seeking to launch an environmental action.

In contrast, South Africa’s approach to standing should make it easier for a litigant to place an environmental issue before the courts – at least theoretically. Prior to the advent of

129 Massachusetts (note 129 above) at 505 footnotes 2–4.
130 Ibid at 520.
134 Ibid at 561–562.
137 Ibid at 847–850.
138 Murombo (note 120 above) at 170.
democracy in 1994, legal standing was only afforded to those whose interest in a matter was ‘sufficient, direct and personal’. This is similar to the injury and causation requirements of standing in the USA. The Constitution, however, explicitly states in s 38 that a broad number of individuals or groups has access to a court. Such persons may act in their own interest, in the interest of others, and even in the public interest. In this way, very broad standing is provided with regards to constitutional rights, such as the environmental rights contained in s 24. Further, s 32(1) of the National Environmental Management Act 107 of 1998 (NEMA) essentially provides everyone with standing to approach a court about an issue related to the Act. In particular, these provisions loosen the requirements regarding the nature of the interest a potential litigant must hold in the matter at hand. Regarding the causation requirement which has caused difficulties in the USA, this might also be less problematic in the South African context since the case of Lee. Here the but-for test, traditionally used to establish factual causation, was softened so that this element of a delict could be proven if it was more probable than not that the defendant caused the harm. While the decision has been criticised for promoting uncertainty in the law of delict, in the environmental sphere it may be a very useful development. Environmental issues such as climate change or biodiversity loss have composite causes, meaning that the strict but-for test – focused as it is on finding one factual cause – may not be satisfied by any one actor contributing to environmental damage. However, the contribution of such an actor may be an important element of the composite cause, and a flexible test is needed to accommodate this.

In summary, therefore, environmental agencies such as the EPA have assisted in mainstreaming environmental issues in other jurisdictions. It is not necessarily the case, however, that such an agency will assist in establishing legal standing for environmental litigation, as was the case in the USA. Further, the test for standing in the USA differs materially from South Africa’s approach, where constitutional rights and the relaxation of common law principles should theoretically promote the cause of environmental action. The fact that they do not only adds to the anomaly of the silence of environmental rights as discussed above. It suggests that standing may not be the primary barrier in the South African context.

B Specialised environmental courts

The second international approach which may be considered is the creation of a specialised environmental court or tribunal (‘ECT’). This is in fact becoming a dominant method of

---

139 Director of Education, Transvaal v McCagie & Others 1918 AD 616, 623.
140 Constitution, s 38(a).
141 Ibid, s 38(b)–(c).
142 Ibid, s 38(d).
143 It should be noted, however, that similarly broad standing does not exist with regard to common or customary law dealing with environmental issues Wildlife Society of Southern Africa & Others v Minister of Environmental Affairs and Tourism of the Republic of South Africa & Others 1996 (3) SA 1095 (Tks) 1106.
144 Lee v Minister for Correctional Services [2012] ZACC 30, 2013 (2) SA 144 (CC)(‘Lee’).
145 Minister of Police v Skosana 1977 (1) SA 31 (A).
146 Lee (note 146 above) at para 55.
addressing environmental issues,148 with 1 200 ECTs existing across 44 countries in 2016, and 1 500 by 1 March 2018.149 One example is the Land and Environment Court of New South Wales in Australia, which has functioned for almost 40 years.150 Another is the Agricultural and Environmental Court in Bolivia (Tribunal Agroambiental).151 A new and relatively successful forum is the National Green Tribunal in India,152 which through a number of decisions has clearly established the principle that environmental damage may violate fundamental rights.153

The preference for ECTs may be related to their potential for centralising environmental expertise,154 so as to address the complexity of the issues mentioned above. Relatedly, they may also allow for a particular environmental issue to be addressed in a focused manner. For example, the Environmental Court in Hermanus, South Africa was established primarily to deal with abalone related offences.155 This court had a very high success rate,156 but it was closed down in 2006 and South Africa currently has no designated ECT.157

The fact that an ECT has already existed in South Africa might simplify the process of setting one up again, but the fact that it closed down so soon after it opened suggests that there may have been barriers to its functioning. While it is not clear why the court closed down, beyond the fact that the then Minister of Justice was given a ‘proposal’,158 possible explanations may be made based on an experience in a similar context. Kenya’s Environment and Land Court has faced challenges in negotiating the resource constraints and competing priorities of a postcolonial African country.159 This is not an impossible barrier, and the court has been generally successful in starting to build a robust environmental jurisprudence.160 However, the Kenyan experience – together with its similarities to South Africa’s own postcolonial position – suggests that the closure of the Hermanus’ Environmental Court may have been at least partly related to resource constraints.

157 Phakathi (ibid).
159 Kaniaru (note 150 above) at 580–581.
If this is the case, it is an argument against the use of specialised environmental courts to address the silence surrounding s 24. A further argument is that such courts may not address the problem as presented above, the environmental blindspot at the Constitutional Court level. An environmental court would likely provide a forum for disputes that are clearly related to s 24, and some of these disputes might reach the Court. I would suggest, however, that an environmental court would not assist where cases are heard in other courts, and their environmental ramifications are not identified. This is what I argue occurred in the three cases discussed above. It is a failing that may not have been significant to the parties to the cases, but it was a lost opportunity to mainstream environmental issues and build a stronger environmental jurisprudence. This resulting gap is a problem for future parties who will face infringements of their environmental rights.

C Existing practices in South Africa

I have suggested above that the failure of the courts to identify environmental infringements may be connected to issues of complexity, conflicting priorities and standing. A possible consequence of these challenges is that environmental issues are often not pleaded before court, even though they might be relevant. This is a difficult practical issue, as it is well established that courts cannot address issues that are not placed before them. This is true even of the Constitutional Court, despite its wide jurisdiction.161 The dynamic has led De Wet to comment that the Court has not yet had an opportunity to engage with s 24 sufficiently.162

However, there is inconsistency in the approach to procedural rules at the Constitutional Court. This can be demonstrated by considering various judgments written by Jafta J, the first of which is his minority judgment in Rivonia Primary School. The case concerned the powers of the provincial education department to instruct a public school to exceed the limits in its admissions policy so as to allow a little girl to attend the school closest to her home.163 It was held that the provincial education department did have these powers, and was right to exercise them, although they did not do this in a manner that was procedurally fair.164 Jafta J (together with Zondo J) concurred with the judgment in part, but also reminded the majority that the parties had not in fact referred to procedural fairness in their pleadings. Technically then, this issue should not have been decided – although it was.

Jafta J appears to set much store by correct judicial procedure. He uses a similar approach in the case of KwaZulu Natal Joint Liaison Committee,165 where he was again in the minority, writing with Mogoeng CJ. Here, the issue related to the payment of subsidies to private schools in KwaZulu Natal for 2009,166 and the majority written by Cameron J held that additional payments were necessary.167 However, Jafta J and Mogoeng CJ pointed to factual gaps in the pleadings, and argued that the Court should have only adjudicated the issue that was properly

161 Lagoonbay (note 18 above) at paras 38–39.
162 de Wet & du Plessis (note 8 above) at 346.
164 Ibid at para 68.
165 KwaZulu Natal Joint Liaison Committee v MEC for Education, KwaZulu Natal & Others [2013] ZACC 10, 2013 (4) SA 262 (CC) (‘KwaZulu Natal Joint Liaison Committee’)
166 Ibid at paras 1–8.
167 Ibid at para 78.
set before it.168 Their minority is clearly a response to Cameron J’s somewhat flexible attitude to procedure. His majority again demonstrates that the Constitutional Court does not always follow its own established procedures to the letter.

Interestingly, however, despite identifying inconsistencies in the Court’s jurisprudence, Jafta J himself has not always adhered to procedural rules. The case of Kirland169 concerned a defective provincial government decision to approve an application for the construction of a private hospital in the Eastern Cape.170 A decision to refuse the application was not upheld by then Acting Superintendent-General, so that the application was improperly granted. Here, the majority (again written by Cameron J) held that the decision should be upheld – despite being unlawful – because government had made no formal application for it to be set aside.171 Jafta J again wrote a minority judgment, but this time he challenged the adherence to procedural rules, arguing that the majority ‘place[d] form way above substance’.172

It appears then that it is not always necessary for an issue to have been pleaded before the Constitutional Court for it to be taken up and considered by its judges. Jafta J has been astute in recognising moments in the jurisprudence where the court has loosened its procedural rules. However, at other moments Jafta J himself has adopted this approach. Boonzaier usefully identified that while Cameron J and Froneman J overlooked procedural rules in KwaZulu Natal Joint Liaison Committee, and Jafta J and Zondo J challenged this approach, the pattern was reversed in Kirland.173

While such inconsistencies may create uncertainty, they may also create spaces for judicial innovation and development. Boonzaier has suggested that this is how the Court’s varying approach to procedure should be understood,174 and the results may be positive. Devenish has praised the Court for bypassing procedure to give effect to constitutional rights in cases such as Rivonia Primary School.175 Couzens has made a similar point regarding the child-focused jurisprudence of the Court in C v Department of Health.176 She prefers the substantive approach of the majority to the Jafta J’s literal approach in the minority.177 Thus, the flexibility in the Court’s approach to procedure may be an opening for the development of new techniques in rights-focused jurisprudence. In particular, it might be an opening for the development of a greater jurisprudence surrounding s 24, despite the many barriers which may prevent an environmental issue being raised in litigation. This would likely be in line with the ‘proactive jurisprudence’ suggested by Devenish,178 or the ‘culture of rights’ advocated by Langa J in S

168 Ibid at paras 183–190.
169 Kirland (note 87 above).
170 Ibid at paras 6–26.
171 Ibid at para 106.
172 Ibid at para 50.
174 Ibid at 26.
176 C & Others v Department of Health and Social Development, Gauteng & Others [2012] ZACC 1, 2012 (2) SA 208 (CC)(‘C v Department of Health’)
178 Devenish (note 177 above) at 499.
Further, it may link well with the well-known concept of transformative constitutionalism, as introduced in the South African context by Klare, and developed more recently and in a broader context by others such as Kibet and Fombad.

If the court is to take on a more proactive role – or is doing so already – this may not be dissimilar to the High Court’s position of upper guardian for children’s rights. This was mentioned above, and means that the High Court is mandated to raise issues that are relevant to the promotion of children’s interests, regardless of whether the parties have done so. I would suggest that a similar institution might be useful in the environmental context – an upper guardian for the environment, perhaps. I have already discussed the similarity of challenges that children and the environment may face in getting to court, as well as being heard. I have also discussed the environmental blindspot, and its parallels to the past – and perhaps ongoing – gender blindspot seen in the courts. An upper guardian might assist in resolving these issues, with a mandate to consider the cases before it through an environmental lens, and identify the environmental implications of these cases when the parties fail to do so. Should an existing court be named upper guardian, this would side-step the challenges facing specialised environmental courts as discussed above. An existing court – the High Court, for example – would hear those cases whose environmental ramifications have not been identified by the parties, and could identify and comment on these. An existing court would also not need to be separately established, so that resource constraints are less likely to be a barrier.

In summary, therefore, it appears that the Constitutional Court at times makes use of its wide jurisdiction to decide issues even where they have not been fully pleaded before it. The Court may be developing new judicial techniques in this way, and these are to be welcomed where they allow for the realisation of constitutional rights. I have suggested that this flexibility may be particularly useful in the advancement of environmental rights, whose relevance may not always be recognised by the parties. In this way, the Court could make use of its existing judicial tools to further develop an environmental jurisprudence, without the necessity of an environmental agency or specialised court. However, in order to encourage this development, a specified institution within the existing framework of courts may be useful. I have suggested that we may draw lessons from children’s rights, and consider the development of a court with the role of upper guardian for environmental rights. In this way, existing judicial tools may be expanded upon for the development of a more extensive jurisprudence surrounding s 24.

VII CONCLUSION

The last time the Constitutional Court engaged with constitutional environmental rights was in Fuel Retailers in 2007. Here, Ngcobo J established sustainable development as a guiding concept for environmental rights in South Africa. This is an idea with much international recognition, but it is also a novel and developing concept, so that its use by Ngcobo J means that a certain degree of uncertainty was inserted into s 24. The flexibility of the concept also

---

182 The Children’s Act, s 45(4).
meant that Ngcobo J could argue largely along the economic lines, while ignoring the social and environmental components of sustainability.

The uncertainty of the precedent set by *Fuel Retailers* may have contributed to the silence of s 24 in more recent years. There has been no other meaningful engagement with the right, and various cases came before the Court in which s 24 might have been relevant, but was not argued. Above, I have considered three cases in more detail, so as to illustrate the opportunities that the Court missed for the development of a sound environmental jurisprudence for the protection of s 24 rights. There are various possible reasons for this, beyond the difficulties related to *Fuel Retailers*, and I have suggested that some of these are related to complexity, competing priorities, and standing. Environmental issues are highly complex, and knowledge of them is quickly evolving, so that courts may be hesitant to engage with environmental rights. Further, environmental issues are closely related to social issues, and in an environment of resource constraints this may mean that a court must focus on the social elements of a system. This is despite the fact that a failure to address environmental challenges in a complex socio-ecological system is likely to mean that social action has limited usefulness. This failure to see the environmental ramifications of a development, be it water supply, mining or the position of a petrol station, may be likened to the gender blindspots that have been observed in the courts in other cases. Another barrier that may prevent environmental issues from reaching the courts is legal standing. Environmental issues have faced this challenge internationally, and it may be likened to the difficulty that children may face in getting to court. In South Africa, there are various systems in place to deal with this challenge and ensure that children have access to the courts.

In seeking to overcome these challenges, this article explored some of the tools that have been used internationally to improve access to court for environmental issues. Environmental agencies are well known to function in some jurisdictions, and these might be thought to represent the environment in judicial matters, and so defeat the challenge of legal standing. However, the well-established EPA in the USA has not necessarily acted in this manner, although it does further environmental interests in many other ways. In the context of South Africa, the challenge of legal standing is not in fact the most significant barrier, as citizens can bring an action based on any violation of constitutional rights, even if it does not affect them directly.

The second international tool considered is the establishment of a specialised environmental court or tribunal. Such courts have been very successful in other jurisdictions. However, it is notable that the specialised environmental court in Kenya has faced challenges related to resource constraints. Similar challenges might be seen in South Africa given the postcolonial positioning that it shares with Kenya. But beyond this, such a court might not be the best forum to address the environmental blindspot. Such a court will only hear explicitly environmental cases. It would not have the opportunity to contribute towards the mainstreaming of environmental issues by identifying cases where the relevance of s 24 has been overlooked.

Therefore, it may be that neither an environmental agency nor a specialised environmental court is the most suitable tool for breaking the silence on s 24. Neither would assist where environmental issues have not been pleaded before South African courts. In such cases, procedural rules would dictate that environmental issues cannot be considered, even if they exist. However, I have argued that the Court is inconsistent in its approach to procedural rules. Although this may inject uncertainty into its jurisprudence, the flexibility may also provide the Court with space to develop new judicial techniques for the realisation of constitutional
rights. In particular, the Court might use this space to develop its environmental rights jurisprudence. A more proactive approach might even be formalised in something akin to the upper guardian role of the High Courts, which currently exists for children’s rights. This would provide the designated court with a clear mandate. If it were the High Court that was named the upper guardian for environmental rights, then this might also extend the Court’s procedural flexibility to a broader range of cases. This flexibility would be appropriate for cases of environmental relevance, where there are various reasons that parties may not identify the environmental consequences that relate to a matter. In this way, there are tools within South Africa’s existing judicial frameworks that may be developed to build a stronger environmental rights jurisprudence. These tools should be recognised and called into greater service, so that environmental rights can be fully realised.
Judicial Independence and the Office of the Chief Justice

C H POWELL

ABSTRACT: This article investigates the extent to which the Office of the Chief Justice (OCJ) promotes the independence of the judiciary in South Africa. Judicial independence is widely understood to be protected by security of tenure, financial independence and administrative independence, three characteristics which are meant to support the judiciary as an institution, as well as the independence of individual judges. However, current jurisprudence and scholarship fail to engage with the relationship between individual and institutional independence, and to identify mechanisms of protection for the institution as such. The factors which have received the most emphasis are the financial independence of the judiciary and the judiciary’s control over its own administration. The article reveals that the OCJ has taken over broad areas of the administration of the judiciary, but questions whether the increased control enjoyed by the leadership of the judiciary has translated into improved control for individual judges. It draws on the legal philosophy of Lon L Fuller to suggest how the independence of individual judges relates to the independence of the institution. In particular, it applies Fuller’s theory of ‘interactional law’ to suggest that a process of mutual engagement is needed within those institutions which have to uphold the rule of law. From this perspective, it appears that the OCJ may not be in a position to protect the institutional independence of the judiciary, because it does not contain the mechanisms to accommodate the input of individual judges on the best conditions for effective and independent work.

KEYWORDS: judicial independence; Fuller; rule of law; interactional law; judicial self-governance; culture of justification.

AUTHOR: BA LLB (UCT) LLM (Humboldt), SJD (Toronto)
Email: Cathleen.Powell@uct.ac.za

ACKNOWLEDGEMENTS: This article was an adventure lasting more than two years, and a whole team of constitutional lawyers got me through to the other side of it. For information on the practice of the judiciary and OCJ as well as information on the Chaskalson/Langa report, I thank Chris Oxtoby, Hugh Corder, Richard Calland, Franny Rabkin and Yvonne van Niekerk. For much-needed comments on my various drafts, I am grateful to Chris Oxtoby, Franny Rabkin and Richard Calland again, the anonymous referees of the CCR, and my long-suffering editors. For research assistance and editing, I thank Jonathan Strug and Amy Sinclair. Thank you, too, to Khomotso Moshikaro for initially suggesting the project.
I  INTRODUCTION

This article explores the extent to which the Office of the Chief Justice (the OCJ) achieves one of its central goals: the constitutionally mandated promotion of judicial independence in South Africa.  

The OCJ is a fairly new creation and, as the discussion below will demonstrate, relatively opaque in its workings. Given this lack of clarity, part II begins with a brief account of the origins of the OCJ, the form that it currently takes, and how its mission is conceived. Part III reviews South African jurisprudence and scholarship on the notion of judicial independence. It offers a brief exegesis on Canadian law on the subject, as that country’s jurisprudence has informed South Africa’s view of what independence means. Part III sets out South Africa’s current approach to judicial independence and identifies some of the theoretical weaknesses in this approach. It covers the various characteristics that are understood to support judicial independence but focuses on the somewhat chimerical concept of institutional independence. While numerous cases emphasise the importance of institutional independence, there is a dearth of coherent theory to explain what specific factors serve to protect the judiciary as an institution as distinct from the factors that are needed to protect individual judges.

Nonetheless, two criteria do emerge from the current literature for institutional independence, and I apply these in part IV to establish an interim ‘score-card’ for the extent to which the OCJ protects judicial independence in South Africa. Part IV’s investigation focuses on the two issues highlighted in South African jurisprudence and scholarship thus far: control over the funds awarded to the judiciary; and control over administration of the judiciary. Applying these criteria leads to an inconclusive result, however, and I go on to argue that the unsatisfactory nature of this result bears out some of the critiques of current conceptions of judicial independence.

Part V then takes a step back to sketch an alternative framework for the concept of judicial independence. The alternative framework is rooted in Lon L Fuller’s theory of the rule of law. I focus, in particular, on Fuller’s account of how a society needs to operate under the rule of law and suggest that the particular process which Fuller considers necessary – what has been called ‘interactional law’ – needs to be found within the judiciary as an institution before the judiciary will be able to uphold the rule of law in its dealings with litigants, other branches of government and the society which it is meant to serve. This discussion connects the rule of law to judicial independence, arguing that we cannot hold the judiciary to be truly independent unless it promotes interactional law in its internal processes and maintains transparent communication with those outside the institution.

To investigate whether the requirements of ‘interactional law’ are met on the ground, I then examine two events within the work of the CJ and the OCJ – the drafting of the Norms and Standards and the Case Flow Management projects – which suggest where the OCJ may be


2 I have deliberately omitted a detailed discussion of the third main aspect of judicial independence, namely, security of tenure. The main reason for this is that the Office of the Chief Justice (‘OCJ’) has a limited role to play in protecting it. Not only is it guaranteed by the Constitution, but judges can be removed only through a process run by the Judicial Service Commission (JSC). As such, security of tenure falls outside the ambit of this article.
failing to promote the rule of law within its own governance structures, and thereby weaken the independence, not only of individual judges, but of the judiciary as an institution.

II THE ORIGINS AND MISSION OF THE OCJ

The OCJ was established in 2010 by a proclamation under the Public Service Act 103 of 1994.3 The decision to establish such an office had been reached by agreement between the then Minister of Justice, Jeff Radebe, and the incumbent Chief Justice, Sandile Ngcobo, but the actual form that the office should take was still undecided. At the request of Chief Justice Ngcobo, former Chief Justices Arthur Chaskalson and Pius Langa headed a ‘Committee on Institutional Models’ to suggest the appropriate form which such an office should take, how it should relate to other branches of government, and what the governance structure of the judiciary as a whole should entail. The report, produced on 22 September 2011,4 was later taken up to a limited extent by the current Chief Justice, Mogoeng Mogoeng. The 2013 document produced by the OCJ on its own establishment sets out the function of the institution as ‘capacitating’ the Chief Justice to execute his functions adequately without relying on the Executive. According to the report, this would allow for reforms which would improve service delivery, address the administrative challenges facing the judiciary, improve access to justice, and promote judicial independence and the doctrine of separation of powers.5

The OCJ presents its own vision as a ‘[a] single, transformed and independent Judicial system that guarantees access to justice for all’6 and describes its mission as being to ‘provide support to the judicial system to ensure effective and efficient court administration services’.7 It undertakes to uphold the values of respect for, and protection of, the Constitution, honesty and integrity, openness and transparency, and professionalism and excellence. These values will ensure ‘accountability of the Judicial branch of the State to the people of South Africa … public confidence in the Judiciary; and respect for the rule of law’.8

There are significant differences between the Chaskalson/Langa vision of the OCJ and the institution as it is constituted today. These will be discussed in more detail below, but two points should be noted at the outset. The first is that Chaskalson and Langa saw the OCJ being set up through three phases. The first was to be the establishment of the ‘OCJ as a national department located within the public service to support the Chief Justice as the head of the Judiciary and the head of the Constitutional Court’ 9 The second phase was meant to set up the OCJ as an independent entity and the third to establish a structure to provide for ‘judicially-led

---

3 Proc 44 GG 335500 of 23 August 2010.
7 Ibid.
8 Ibid.
court administration’. The three-stage vision is still contained in the 2013 report, but, as will be noted below, the OCJ has not yet moved beyond phase one.

The second major difference concerns the governance structure of the judiciary and the OCJ’s role within it. The Chaskalson/Langa report recommended the establishment of a ‘Judicial Council’, which would have formed a second governing body within the judiciary. Amongst its other functions, it would have exercised oversight over the OCJ. Although consisting mostly of judges, it would also have included a broader cross-section of representatives of the legal profession, academia, Parliament and the executive. However, this vision has not been realised. The Superior Courts Act 10 of 2013 failed to mention a Judicial Council at all. Mogoeng CJ’s final recommendation for the OCJ, currently before the executive, retained the notion of a Judicial Council, but reduced it to the Heads of Court. This is a pre-existing institution described by the Chaskalson/Langa report as an ‘informal body consisting of the heads of all superior courts that for many years have been a sounding board for the Chief Justice’.

III INSTITUTIONAL INDEPENDENCE IN CASE LAW AND COMMENTARIES

Determining whether the OCJ has the capacity to protect judicial independence, and whether it is doing so in fact, requires a clear conception of what judicial independence means. The aim of this part is both to give as detailed a picture as possible of how judicial independence is understood in the South African discourse and to point to some flaws in the theory underlying this conception.

In 2002, the Constitutional Court had to assess the independence of the magistracy in the case of Van Rooyen. It relied on the Canadian case of Valente v The Queen. Valente attributed an internal, or subjective, as well as an external, or objective, facet to independence. This conception of independence requires that the judicial officers themselves are impartial in their state of mind and also that there is ‘a status or relationship to others, particularly to the executive

---

10 Ibid.
11 See the discussion in Part IV C.
12 Chaskalson/Langa report (note 4 above) at paras 4.1.1–4.1.2.
13 Ibid at para 4.2.6.
14 F Rabkin ‘Judiciary on Trial: Draft Rules Aimed at Tackling Inefficiencies in the Judiciary Have Caused a Stir and Have Been Slated by Critics as Inimical to Judicial Independence’ Financial Mail (7 March 2014).
15 Chaskalson/Langa report (note 4 above) at para 3.2.1.6. The report pointed out that this body has no formal powers, no infrastructure and no budget. It usually meets two or three times a year to debate matters of concern to the judiciary, to advance their interests in dealings with the executive which has been responsible for all aspects of court administration, and in formulating policies for the judiciary. Interestingly, while implying that the position of the Heads of Court has changed where formal powers, infrastructure and budget are concerned, the recommendation currently before the executive repeats the Chaskalson/Langa description of how often it meets and what it does. See Committee on Institutional Models Report Capacitating the Office of the Chief Justice and Laying Foundations for Judicial Independence: The next Frontier in our Constitutional Democracy: Judicial Independence (subsequent version) (undated, subsequent to Chaskalson/Langa report (note 4 above) September 2011) (on file with the author) (Later Version Committee on Institutional Models Report) at para 4.2.1.6. For further discussion of the powers and functions of the Heads of Court under the 2013 Act, see the discussion below at part VI B.
16 S & Others v Van Rooyen & Others (General Council of the Bar of South Africa Intervening) [2002] ZACC 8, 2002 (5) SA 246 (CC) (‘Van Rooyen’).
17 Valente v The Queen (1986) 24 DLR (4th) 161 (SCC) (‘Valente’).
branch of government, that rests on objective conditions or guarantees.\textsuperscript{18} \textit{Valente} identified three specific elements which together guarantee external independence, namely, security of tenure, financial security and the independence of the institution to which the judicial officers belong.\textsuperscript{19} The court in \textit{Valente} saw security of tenure and financial security together as safeguards of the individual judge’s freedom from outside influence, whereas institutional independence protects the judges as a collective. Distinguishing broadly between individual and collective independence, the court emphasised that the latter aspect was indispensable:

The relationship between these two aspects of judicial independence is that an individual judge may enjoy the essential conditions of judicial independence but if the court over which he or she presides is not independent of the other branches of government, in what is essential to its function, he or she cannot be said to be an independent tribunal.\textsuperscript{20}

In \textit{Valente}, collective, institutional independence related to the control which judges have over the job they have to do. The court described it as the facet of independence concerned with ‘matters of administration bearing directly on the exercise of [the court’s] judicial function’.\textsuperscript{21} However, it then distinguished further between the \textit{adjudicative} and the purely \textit{administrative} judicial functions, according decisive importance only to the adjudicative aspect. This aspect, according to \textit{Valente}, involves the ‘assignment of judges, sittings of the court and court lists – as well as related matters of allocation of court-rooms and direction of the administrative staff engaged in carrying out these functions’.\textsuperscript{22} The purely administrative aspect of the judicial function, on the other hand, was seen in terms of various financial and discretionary benefits enjoyed by the judges, the control exercised by the executive over these benefits and the recruitment and control over support staff within the court.\textsuperscript{23}

In \textit{Van Rooyen}, the Court took its cue from \textit{Valente}, which meant it distinguished between adjudicative aspects of the judicial function and other administration, and required institutional independence only where the former was concerned. But the judgment does not explain what these adjudicative aspects are. As in \textit{Valente}, the relevant aspects were defined in general terms as administrative decisions bearing ‘directly and immediately on the existence of the judicial function’\textsuperscript{24} and as ‘matters that relate directly to the exercise of the judicial function’.\textsuperscript{25} There has been no guidance from subsequent case law on the dividing line between administration which is essential to the judicial function and ‘merely administrative’ functions that can be handled by the executive.

A later Canadian case critiqued \textit{Valente} for not clearly defining institutional (collective) independence.\textsuperscript{26} In \textit{Reference re Public Sector Pay Reduction Act},\textsuperscript{27} Lamer CJC\textsuperscript{28} suggested that

\begin{itemize}
  \item \textsuperscript{18} Ibid at 170 (as cited in \textit{Van Rooyen} (note 16 above) at footnote 22.)
  \item \textsuperscript{19} Ibid at 176–190 (as cited in \textit{Van Rooyen} (note 16 above) at para 29.)
  \item \textsuperscript{20} Ibid at 171.
  \item \textsuperscript{21} Ibid at 187.
  \item \textsuperscript{22} Ibid at 188.
  \item \textsuperscript{23} Ibid at 188–190.
  \item \textsuperscript{24} \textit{Van Rooyen} (note 16 above) at para 29.
  \item \textsuperscript{25} Ibid.
  \item \textsuperscript{27} Ibid.
  \item \textsuperscript{28} The acronym stands for ‘Chief Justice of Canada’.
\end{itemize}
Valente’s theoretical framework was inadequate and proposed that security of tenure, financial security and what he termed administrative independence may each protect both individual and institutional independence. Individual and institutional independence, in Lamer CJC’s view, were dimensions of an independence protected by its ‘core characteristics’ of security of tenure, financial security and administrative independence.29 Lamer CJC stated that these three characteristics –

[c]ome together to constitute judicial independence. By contrast, the dimensions of judicial independence indicate which entity – the individual judge or the court or tribunal to which he or she belongs – is protected by a particular core characteristic.30

Having noted that Valente does not provide a specific vision of institutional independence,31 Lamer CJC suggested his own version in Reference re: Public Sector Pay Reduction Act. The case before him dealt exclusively with the remuneration of judges, but Lamer framed the issue of judicial independence more broadly by linking it to the doctrine of separation of powers.32 The ‘core aspect’ in issue here was, clearly, financial independence, and Lamer CJC had to find a way in which the judicial institution as such could be protected in this regard. He held that the doctrine of separation of powers requires that the courts both be, and appear to be, free from political interference through economic manipulation, and that they not become involved in the ‘politics of remuneration’.33 The only means to achieve this, in his opinion, was by interposing an independent body between the judiciary and other branches of government.34 Although Lamer CJC did not insist that the body’s views be binding on the Executive, he did require that the body itself be independent, effective and objective.35 Its members themselves must have security of tenure36 and the body should not be controlled by any of the branches of government.37

In South Africa, the ideas that emerge from policy frameworks, proposals and discussions echo those underlying Van Rooyen and the Canadian jurisprudence. The trio of security of tenure, financial independence and administrative independence are emphasised, there is a strong stand against any form of pressure being brought to bear against individual judges, and the importance of the judiciary’s independence is emphasised in the abstract.38 But this is a far cry from explaining quite how judges as a collective, or the judiciary as an institution, are to be protected. As noted above, the Court largely took its cue from Valente and limited institutional independence specifically to adjudicative aspects of the judicial function, without defining what

29 Reference re Public Sector Pay Reduction Act (note 26 above) at para 118.
30 Ibid at para 119.
31 Ibid at para 123.
32 Ibid at para 125.
33 Ibid at para 131.
34 Ibid at paras 133 and 166.
35 Ibid at paras 170–175.
36 Ibid at para 171.
37 Ibid at para 172.
these are. In various proposals by, among others, the Department of Justice, Ngcobo J, the Chaskalson/Langa proposal for an Office of the Chief Justice, and UCT’s Democratic Governance and Rights Unit, there is still a sense that only some aspects of administration ‘bear directly on the exercise of the judicial function’ but there is little clarity on where the boundaries of this administration lie. For his part, Ngcobo J suggested that case-flow management and records, statistics and information management is so closely connected to the judicial function that it can be handled only by the judiciary. However, at this stage, he did not consider the executive handling of budgetary and financial management, space and equipment management and other administrative functions as ‘incompatible with judicial independence’. He noted, in addition, that the government has to provide buildings and funding for the judicial project.

From a practical perspective, the exact dividing line between ‘adjudicative’ and ‘purely administrative’ functions may appear almost irrelevant. As will be seen in part IV below, the OCJ has in fact taken on such a wide swathe of administrative functions that it is easy to argue that the ‘purely adjudicative’ aspects of that administration must be under its control as well. But, from another perspective, this wrangling about the distinction between adjudicative and purely administrative administration is important, because it demonstrates how unclear the theoretical basis is for a coherent picture of institutional independence. A good example can be found in Valente. As seen above, Valente equated institutional independence with administrative independence, while seeing financial security and security of tenure as guarantees of the independence of individual judges. But this leads to anomalies. The case appears to slot control over administrative personnel into both the adjudicative and administrative aspect of institutional independence. It categorises as adjudicative the direction of the administrative staff engaged in carrying out the court’s functions and as administrative the control over support staff within the court. Furthermore, the case sees discretionary benefits as an issue under the peripheral, ‘purely administrative’ aspect of institutional independence. But discretionary benefits affect the individual, not the institution. Valente appears to be ignoring its own main distinction between the independence of the individual and the independence of the institution.

Lamer CJC’s subsequent refinement of the Valente idea accepts that the ‘core characteristics’ of security of tenure, financial security and administrative control can affect both individual judges and the judiciary as an institution. Thus ‘institutional independence’ consists of something more than mere administrative independence. But this refinement has not assisted the South African discussion to flesh out what the ‘value added’ of institutional independence might be. Thus, in Van Rooyen, the Court did suggest additional conceptions of institutional

39 Van Rooyen (note 16 above) at para 29.
40 Department of Justice Policy Framework (2010)(on file with author) (‘DOJ Policy Framework’).
42 Chaskalson/Langa report (note 4 above).
43 Much of this research was undertaken to support the Chaskalson/Langa report (note 4 above).
44 DOJ Policy Framework (note 40 above) at para 71 and Ngcobo (note 41 above).
45 Ngcobo (note 41 above).
46 DOJ Policy Framework (note 40 above) at paras 78 and 80; Ngcobo (note 41 above).
47 Valente (note 17 above) at 188.
48 DOJ Policy Framework (note 40 above) at para 71.
independence over and above administrative independence. One of these is that institutional independence is the independence of the courts from other arms of government.49 But it did not build on this conception because it provided no tools to ‘protect courts and judicial officers from external interference’.50 This is because it saw appeal and review procedures by other courts as a mechanism of institutional independence of the magistracy.51 This is problematic because, even in the case of the magistracy, it provides merely a remedy after the interference has taken place rather than structural protection against the interference taking place at all. But, transferred to the judiciary as a whole, the idea that judicial review can protect the independence of the judiciary is meaningless. Its reliance on the judicial hierarchy for protection means that there is no remedy if it is the top court itself which lacks independence.52 In effect, calling on judicial review to protect judicial independence requires the judiciary to provide the very guarantees without which it cannot function in the first place, a circular and self-defeating notion.

For its part, the OCJ has not engaged with the conceptual problem that emerges from this survey. Instead, it focuses strongly on ‘a system of court administration based in the judiciary’ when it envisages judicial independence,53 and, as we will see below, it has taken over control of large sections of this administration.

IV INTERIM SCORE-CARD: THE OCJ FROM THE PERSPECTIVE OF BUDGET NEGOTIATIONS AND ADMINISTRATIVE CONTROL

At this stage, we have extracted two main markers of independence from the confusion of case law and legal discussion: non-political control over the budget of the judiciary, and judicial control over the administration of the courts. I have suggested that these markers are inadequate and unclear, but in this part I nonetheless evaluate the achievements of the OCJ against them in detail. The in-depth discussion, will, I hope, achieve three main objectives. First, it will suggest something of an interim assessment of the OCJ’s achievements in promoting judicial independence as hitherto conceived, although the evidence does not lead to a clear result. Second, and more importantly, it will underscore my theoretical critique of the current independence criteria by demonstrating the vague and inconclusive results they achieve when applied to facts on the ground. Thirdly, the factual description set out here will serve as a starting point for an alternative evaluation of the OCJ’s role in protecting institutional independence, one which is explored in more detail in part V.

A Budget

The Reference re Public Sector Pay Reduction Act case suggests that a separate body be set up to mediate between the judiciary and the government to address the remuneration of judges. The case does not require that this separate body have binding powers; merely that its members

49 Van Rooyen (note 16 above) at paras 18–19.
50 Ibid at para 19.
51 Ibid at para 24.
52 Ibid at para 18.
53 I am grateful to one of my anonymous reviewers for this point.
themselves have sufficient independence from the other branches of government.\textsuperscript{55} Later Canadian case law requires that the government justify departures from the recommendations of this body.\textsuperscript{56}

The financing model proposed in the Langa/Chaskalson report would have had a separate committee within Parliament to which the OCJ would have accounted directly and to which it would have made budgetary requests. This committee would then have handled financial and budget requests in much the same way as the Auditor-General’s finances are handled. That was, however, planned for phase two of the development of the OCJ.\textsuperscript{57} It has not been adopted by the current model. The place of the OCJ within the executive structures is clearly demonstrated by the procedure whereby its budget is determined. Two bodies play a role in assisting the President to determine judges’ salaries: the Independent Commission for the Remuneration of Public Office Bearers\textsuperscript{58} and the OCJ itself. The Independent Commission for the Remuneration of Public Office Bearers draws up recommendations which the President considers when he determines the salaries, allowances and benefits of judges.\textsuperscript{59} However, the President is not bound by the recommendations. The OCJ plays a role in the determination of salaries to the extent that the Chief Justice is consulted by the Commission during its deliberations.\textsuperscript{60}

Where the rest of the budget for the judiciary is concerned, s 54 of the Superior Courts Act 10 of 2013 requires the Minister to consider requests for funds needed for the functioning and administration of the Superior Courts, as determined by the Chief Justice in consultation with the Heads of Court. The Minister must consider these requests ‘in the manner prescribed for the budgetary processes in the departments of state’. Since 2015, the OCJ budget request has


\textsuperscript{56} The Supreme Court of Canada has held that governments have a constitutional duty to use an independent, effective and objective body for recommendations on salary reductions, increases or freezes for judges. Furthermore, if these recommendations are ignored, then that decision must be justified, if necessary in a court of law, on the basis of a simple rationality test. \textit{Mackin v New Brunswick (Minister of Finance) Rice v New Brunswick} (note 55 above).

\textsuperscript{57} Chaskalson/Langa report (note 4 above) at para 3.2.1.3.

\textsuperscript{58} The Independent Commission for the Remuneration of Public Office Bearers was established in terms of Constitution s 219 and the Independent Commission for the Remuneration of Public Office Bearers Act 92 of 1997.

\textsuperscript{59} Judges Remuneration and Conditions of Employment Act 47 of 2001, s 2.

\textsuperscript{60} \textit{Association of Regional Magistrates of Southern Africa v President of the Republic of South Africa & Others} [2013] ZACC 13, 2013 (7) BCLR 762 (CC) at para 61 (The Constitutional Court rejected a claim that broader consultation was required, holding that the Chief Justice ‘represents the entire Judiciary in such consultations’).
been presented as a separate vote to Parliament. This may have largely symbolic significance, as there is little indication that the separation of the OCJ budget from the two other budgets presented by the Department of Justice and Constitutional Development has reduced the influence of the executive over the budget which is awarded. The budget debates, portfolio committee meetings and reports by the Secretary-General of the OCJ show that, while the OCJ puts its own proposal forward, the final product requires the buy-in of the Ministry of Justice and Constitutional Development and the Treasury, and it is the Minister himself who presents the OCJ budget to Parliament.

In the OCJ’s own description, it presents its case for funds to the National Treasury ‘as with any other [government] Department’. The Secretary-General explained the process in the following terms:

Treasury requested each department to identify the impact that the department would make with the funding that it received. Any additional funds that a department required, had to come from the funding of [its] cluster or functional group. The Minister was ultimately accountable for all funding provided to the departments for which he was responsible.

It is thus fairly clear that the executive currently retains a strong grip on the allocation of funds to the judiciary, which suggests that the OCJ has not, at this stage, helped to improve the independence of the judiciary in this respect. The judiciary’s functioning is affected on an ongoing basis by a shortage of resources and it is not clear what difference the presence of the OCJ has made to this long-standing problem. Indeed, it is possible that the OCJ is taking up money needed elsewhere.

---


62 These are the DOJ’s own budget, and that of Correctional Services. See, for example, Committee Report on Department of Justice, Correctional Services & Office of Chief Justice Budgets (16 May 2017), available at https://pmg.org.za/committee-meeting/24368/.

63 Note, however, that the OCJ remains within the DOJ. As a result, the Minister of Justice who presents the OCJ budget to Parliament. See the Address by Minister TM Masutha, MP (Adv) at the occasion of the budget debate of the Office of the Chief Justice, National Assembly, Cape Town (17 May 2017), available at http://www.gov.za/speeches/minister-michael-masutha-office-chief-justice-budget-vote-201718-17-may-2017-0000 (‘Minister Masutha Budget Debate Address May 2017). Note further that the magistracy is still falls under the DOJ.

64 In his budget speech of May 2017, the Minister of Justice notes the ‘the difficult decisions that the National Treasury needs to make to divide the limited resources against competing priorities of the government’. Minister Masutha budget debate address May 2017 (note 63 above).

65 Minister Masutha Budget Debate Address May 2017 (note 63 above).


67 Ibid.

B Control over administration

In terms of administration, the OCJ appears to have moved well beyond the goals set out in the first phase of the original three-phase plan. The Superior Courts Act 10 of 2013 sets up a management structure for the judiciary with the Chief Justice at its apex. This was followed by the transfer of a broad swathe of functions and personnel to the OCJ in the 2015/2016 year.

The Superior Courts Act 10 of 2013 empowers the Chief Justice to guide, oversee and monitor the administration of the courts in a section headed ‘Judicial management of judicial functions’. The ‘judicial functions’ over which the Chief Justice is given control includes the determination of sittings of the courts, the assignment of judges to the sittings, the assignment of cases and other judicial duties, procedures with respect to case flow management, recesses and the finalisation of matters before a judicial officer.69 The OCJ Annual Report for 2015/2016 notes a further transfer of administration of superior courts to the OCJ from the Department of Justice (DOJ) in March 2015, in terms of the Public Service Act 103 of 1994. The functions and personnel included in this transfer are extensive. They cover judicial support services, case flow management, court operations, language services, the execution of ‘quasi-judicial’ agencies, service level contracts, editing documents and translations, ensuring equal access to all official languages, braille and sign language, procurement, data collection, the forensic unit, public education and communication, and management of a host of units now transferred to the OCJ, such as human resources, finance and security. The personnel involved in these functions, and now assigned to the OCJ rather than the DOJ, include court managers, the registrar, law researchers, court interpreters, support and food service aid, judges’ secretaries, management of the units transferred to the OCJ, IT co-ordinators, administration officers and clerks, statistical officers, vetting investigators and forensic auditors.70 This swathe of functions and personnel gives the OCJ full reign over a much wider area of administration than that suggested as ‘essential to the judicial function’ in the discussion above. Recall that Ngcobo J identified only case-flow management and records, statistics and information management as falling within this category essentially connected to the judicial function. He did not consider the executive handling of budgetary and financial management, space and equipment management and other administrative functions as ‘incompatible with judicial independence’.71 He would not have required that the OCJ take control over the work of financial management and procurement. The assignment of administrative functions to the judiciary also goes well beyond what is required by the Valente test. As noted above, Valente saw the recruitment and control over support staff within the court as purely administrative (not ‘adjudicative’).72 Thus Valente would not have required that the judiciary have control over food services or even human resources management.

One of the ways in which the Chief Justice acted on the powers conferred on him by the Act was by issuing a notice, gazetted in February 2014.73 The notice sets out a clear hierarchy of responsibility and control for the judiciary and prescribes a set of ‘norms and standards’ for its operation. Seen as a pyramid, the hierarchy sets the Magistrates’ Courts at its base. These

---

69 The Superior Courts Act 10 of 2013, s 8(6).
70 Letter to the Minister of Justice, Michael Masutha, from Acting Minister for Public Service and Administration, Nathi Mthetwa (31 March 2015) (on file with author, see explanation note 141 below).
71 DOJ Policy Framework (note 40 above) at paras 78 and 80, Ngcobo (note 41 above).
72 Valente (note 17 above) at 188–190.
73 GN 147 GG 37390 of 28 February 2014.
courts are the responsibility of the Heads of Court of each division into which they fall. The accountability structure is that the lower levels of the hierarchy have to account to the higher levels. The Heads of the Regional and Administrative Magistrates’ Regions have to account for their management of these Regions to the relevant Judge President. The President of the SCA and the Judge President of each Division have to account to the Chief Justice for their management of the judicial responsibilities allocated to them. But there is no regulation of how these functions are to be managed or how the persons assigned to monitor the judicial officers’ fulfilment of their responsibilities are to do so.

The ‘Norms and Standards’ section is divided into ‘objectives’, ‘core values’, ‘management of judicial functions’, ‘monitoring and implementation’ and the ‘norms and standards’ themselves. They provide extensive detail about how courts should operate. Once again, the focus is on what the judges should be doing, with scant attention to the rights and expectations which the judges themselves may have of the institution and its management. The stated ‘objectives’ are set out in the following terms:

[to] achieve the enhancement of access to quality justice for all; to affirm the dignity of all users of the court system and to ensure the effective, efficient and expeditious adjudication and resolution of all disputes through the courts, where applicable. These objectives can only be attained through the commitment and co-operation of all Judicial Officers in keeping with their oath or solemn affirmation to uphold and protect the Constitution and the human rights entrenched in it and to deliver justice to all persons alike without fear, favour or prejudice in accordance with the Constitution and the law.

Under the heading, ‘Norms’, the Notice emphasises the duties of each judge to act efficiently, effectively, expeditiously, courteously and respectfully towards the public, and collegially towards their colleagues. The Heads of Court are encouraged to ‘take all necessary initiatives to ensure a thriving normative and standardised culture of leadership’ and to engender an ‘open and transparent policy of communication both internally and externally’. The section on ‘Standards’ includes provisions on the numbers of hours the trial courts should sit per day and time limits for the delivery of judgments.

Finally, the ‘Norms and Standards’ paragraph sets up a Judicial Case Flow Management Report to be co-ordinated by National and Provincial ‘Efficiency Enhancement’ Committees (the NEEC’s and PEEC’s), the former headed by the Heads of Court and reporting directly to the Chief Justice. This initiative sits at the heart of the OCJ’s current preoccupation with improving the efficiency of the courts. Its three goals of 2015 continue to drive its efforts: capacitating the OCJ (a goal centred on improving the human resources of the office); supporting the Chief Justice in his functions as head of the judiciary; and, finally, rendering...
JUDICIAL INDEPENDENCE AND THE OFFICE OF THE CHIEF JUSTICE

‘effective and efficient administration and technical support to the Superior Courts’. The focus is to ensure that the judges do their jobs and deliver the necessary services to the public. This initiative has been supported by the allocation of funds to a case-flow management programme and to the modernisation of the court system to allow for e-filing.

C Interim assessment: independence and judicial governance

Applied to the facts set out in part IV B, we see that the criteria set out in part III above provide, at best, an ambivalent, and, at worst, a meaningless result. The judiciary seems to score very poorly on one criterion (financial independence) while almost overachieving on the other (administration), as it now has control over much of the administration which the literature on judicial independence would leave in executive hands. The current vision of judicial independence does not help us to balance or interpret these widely differing scores.

The criteria isolated in part III also do not equip us to evaluate much of the structural development – or lack thereof – of the OCJ. First, they do not give us a way of evaluating the fact that the original three-phase plan for the OCJ is not being realised. Most importantly, the OCJ is still a government department (planned for phase one), and has not become an independent entity (planned for phase two). This would appear to be of central relevance to an assessment of the judiciary’s institutional independence, and yet the criteria set out above do not equip us to engage with this issue.

This lacuna in our theory – the inability of scholarship to engage with what makes the judiciary independent as an institution – has led to some incoherent and self-contradictory stances. I have already noted the Court emphasised institutional independence in Van Rooyen while it offered no protections unique to the judiciary as an institution. More recently, an online publication of the OCJ provides two diametrically opposed statements about the need for the OCJ to be an independent entity. The publication, ‘2030 Vision of the Judiciary’, consists of two pieces by the current Chief Justice, Mogoeng Mogoeng. In the first, he states that he is ‘convinced’ that phase two (the phase in which the OCJ becomes an independent entity) is unnecessary, although he provides no reasons for his position. Instead, he simply recommends proceeding straight to phase three without attaining what the original plan


83 OCJ Annual Performance Plan 2018/2019 (note 1 above) 12. See also the report by the Secretary-General of the OCJ to the parliamentary portfolio committee on its Annual Performance Plan for 2016 (6 April 2016), available at https://pmg.org.za/committee-meeting/22279/.


85 M Mogoeng ‘2030 Vision for the Judiciary (A Contribution to the National Development Plan)’, available at https://www.judiciary.org.za/index.php/documents/publications/category/55-judiciary-publications?download=217:2030-vision-for-the-judiciary, 18. The date of publication of this document is unclear. However, it appears that the work was initially released on June 2011. It also appears that a second iteration supplanted the first publication on 25 April 2013. Ibid at 21.
seemed to consider integral to institutional independence. He reports that this approach has the 
support of the executive. In the second piece of this online publication, however, Mogoeng 
CJ sets out the original three-phase plan as though it will still be realised in full, a plan which 
is, in fact, still reflected in the recommendation currently before the executive. We do not 
know what is happening with the original three-phase plan, and we do not know what should 
be happening because we lack the conceptual tools to specify what institutional independence 
really means.

Instead, institutional independence seems to have been conflated with (and reduced to) the 
notion of administrative independence. But the criterion of administrative independence is 
also inadequately realised. As we have seen, the main debate has centred on quite what level of 
administrative control has to be transferred from the executive to the judiciary before we can 
say the judiciary is independent. In the case of the OCJ, far more control has been transferred 
than most of the theoretical discussion required. Perhaps this is the main reason why Mogoeng 
CJ views judicial independence as all but attained:

[What] the OCJ has achieved – [is] to lay a very solid foundation for Judicial self-governance, the 
only remaining barrier to the attainment of complete Judicial independence.

However, the discussion thus far has suggested that we lack the analytical tools to determine 
whether administrative control over the judiciary by the OCJ necessarily equates to judicial 
independence. This is because the current conception of judicial independence has nothing 
to say about how the judiciary should exercise control over its own administration. It equates 
judicial independence with ‘self-governance’, but provides no guidelines on what that 
governance should look like. But, as Hassen Ebrahim warns, we need these ‘governance’ 
guidelines to establish whether the OCJ’s control of the judiciary is the same thing as judicial 
independence:

[Any] strategy that seeks to assume judicial control over the court administration system using the 
platform of the OCJ would merely be replacing the Minister with the Chief Justice as executive 
head over the administration, and would not address the demand to allow judicial officers greater 
control over the administration of their own courts.

Ebrahim’s warning suggests that individual judicial independence is a prerequisite for 
institutional judicial independence. In his view, judges (at every level) need control over their 
own courts before we can speak of judicial independence. Because the OCJ controls the 
judiciary, individual judges therefore need to have a voice within the governance structures 
of the OCJ itself. On this line of thinking, a particular form of judicial self-governance – one 
that allows for the participation of all the judges in the system – is a prerequisite for judicial 
independence.

What, then, is the connection between individual judicial independence and institutional 
judicial independence, and what (if any) are the requirements for a judiciary truly to be 
considered ‘self-governing’? The literature covered in part III does not cover these issues, as it 
does not establish clearly what institutional independence is, nor does it discuss what the link

86 Ibid.
87 M Mogoeng ‘The Implications of the Office of the Chief Justice for Constitutional Democracy in South Africa’ 
(note 85 above) at 21, 28.
88 Committee on Institutional Models Report (note 15 above) at 3–4 and at paras 2.2.3 and 4.1.
89 Mogoeng (note 85 above) at 21, 39.
90 H Ebrahim ‘Governance and Administration of the Judicial System’ (note 54 above) at 114.
between individual and institutional independence might be. But the implication is that there is no link. For example, by limiting individual judicial independence to the security of tenure and financial security of the judges, Valente equates individual judicial independence with the absence of pressure on a judge to decide cases in a particular way. The idea that a judge needs some voice in the governance of the court system before either the individual or the institution can be considered ‘independent’ is not canvassed in Valente or the other sources in part III.

We need, therefore, to take a step back and explore the design requirements of a judiciary which is ‘governing itself’. For this, the legal philosopher Lon L Fuller suggests a fresh way to look at what the rule of law itself might require of the form of judicial self-governance.

The rule of law is self-evidently central to constitutional democracy, and is acknowledged as such by the OCJ for its role in ensuring accountability. However, it is unclear quite how the OCJ understands the term ‘rule of law’ in its reports, other than that the judiciary itself should be ‘handing down’ law and that its determination of the law should be respected by other branches. If we follow Fuller’s approach, however, we will see that, under the rule of law, law has a role to play in designing the judicial institution itself, not just in locating the judiciary in a specific position with respect to the other branches of government.

V A THEORETICAL FRAMEWORK FOR JUDICIAL INDEPENDENCE: LAW IN THE PROCESSES OF THE OCJ

One of Fuller’s innovations was to see law as a process and not just a product. It is, in other words, not just the rules and institutions that govern a society, but the process whereby these rules and institutions are formed. If a society is functioning under the rule of law, a particular process will form and maintain its rules and institutions; a process that is ‘dependent upon the mutual generative activity and acceptance of the governing and the governed’. ‘Mutual generative activity’ relies on ongoing communication and deliberation between the various units of society. The communication needs to include ‘a mode of interaction between actors based on the logic of arguing, that is, of convincing each other to change their causal or principled beliefs in order to reach a reasoned consensus’. If communicative action is understood as central to law, then law becomes a process which accompanies and interfaces with the democratic project. Transferred to the judiciary, Fuller’s ‘interactional theory’ of law would mean that the institution would not just ‘hand down’ law, but be constituted and maintained by it.

The OCJ’s governance structure, to the extent that it is set out formally, is based on a top-down hierarchy, through which the leadership seemingly ensures that the judicial officers remain ‘accountable’ to the public. But it is not designed to make the leadership of the judiciary accountable to its members, or responsive to their views and concerns. In Lon Fuller’s

---

91 Valente (note 17 above) at 176–190, Van Rooyen (note 16 above) at para 29.
92 OCJ Strategic Plan 2015/16 – 2018/19 (note 6 above) at 10; OCJ Annual Performance Plan 2018/2019 (note 1 above) at 12.
96 Brunnée & Toope (note 93 above). See also L Fuller The Morality of Law (1969) 221.
terms, such a structure constitutes ‘managerial direction’ rather than law. In his analysis, the former entails a one-way projection of authority, while the latter is built on a ‘relatively stable reciprocity for expectations between lawgiver and subject’. In Fuller’s view, managerial direction functions for the benefit of the superior, whereas law serves the needs of the subjects themselves and of broader society:

The directives issued in a managerial context are **applied** by the subordinate in order to serve a purpose set by his superior. The law-abiding citizen, on the other hand, does not **apply** legal rules to serve specific ends set by the lawgiver, but rather **follows** them in the conduct of his own legal affairs, the interests he is presumed to serve in following legal rules being those of society generally.

My use of this passage might seem to punt a rather extreme version of the critique of the OCJ as an ‘empire-building’ project. I need to clarify that I am not presenting a spectacle of ‘Emperor Mogoeng’ turning all judges into his minions to further his own, personal benefit. Instead, I am pointing out the likelihood that, without interactional law, he will use the OCJ as a vehicle to punt his own, subjective vision of how the judiciary should be operating within a constitutional democracy. The problem is that the very notion of what the judiciary is and what it needs to be independent is reduced to what the wielder of power within the institution thinks in this regard. And, no matter how benign the Chief Justice’s intention, such a system reduces the judiciary to an extension of the Chief Justice. Judges lose their autonomy and independence when the institution as a whole no longer operates under law.

I would thus argue that the loss of autonomy and independence for individual judges impacts on the ability of the institution as such to hand down law, and thereby reduces the independence of the institution itself. It is irrelevant that the undue influence may be coming from a particular judicial officer rather than an outside person or institution. The judiciary **as such** is hindered in its task of developing and maintaining law. Under Fuller’s rule of law theory, then, individual and institutional judicial independence are intertwined because the rule of law requires a particular form of ‘self-governance’.

Lon Fuller’s ‘interactional theory’ of law – that is, his theory on the process of interaction required by the rule of law – is regarded by current legal theorists as his most fruitful insight. It also resonates with a concept introduced specifically to the South African context as the country was preparing to transition to a post-apartheid legal culture. This concept, a ‘culture of justification’, was coined by Etienne Mureinik to describe the kind of change which South Africa needed to achieve in its move from an apartheid state to a democracy. In contrast to

---

97 Fuller (note 96 above) at 209.
98 Ibid at 207–208.
99 Rabkin (note 14 above).
100 In an interview in 2013, Chief Justice Mogoeng denied that he had power over judges on the basis that he could not fire them (Rabkin (note 14 above)). This argument ignores the fact that judges do not merely want jobs; they want to serve on specific benches or hear particular matters, be promoted to higher courts or enjoy better working conditions. There are, in other words, more benefits at stake than employment as such. The Chief Justice’s influence over the distribution of some of these benefits is by no means transparent. This point is discussed in more detail below.
101 Brunnée & Toope (note 93 above) at 49.
a culture of authority, which had flourished under apartheid, a culture of justification requires that ‘every exercise of power is expected to be justified’, and ‘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’.103 These two perspectives of Etienne Mureinik and Lon Fuller differ slightly. Mureinik focuses on the legal culture required by democracy, but he describes it in terms (justification and, particularly, interaction) which Fuller depicts as quintessential to law itself. Fuller, in effect, incorporates certain features we generally associate with democracy into the very essence of law. The reason that Fuller placed within law some of the characteristics that might be seen as ‘merely’ democratic is that he saw them as indispensable to creating ‘fidelity to law’ in its subjects. In particular, he argued that the ability to reason with norms creates a sense of obligation within law’s subjects to comply with the law without the need for any coercive state mechanisms to enforce it.104

Although they ground the requirement for justification and interaction in slightly different theoretical sources, Fuller’s and Mureinik’s views are not in conflict within a constitutional democracy. Instead, they complement each other and inform our attempts to understand how law requires justification and accountability from those in power. I suggest that Fuller’s theory of interactional law applies within the institutions which exercise power, particularly the institution which claims expressly to be interpreting and developing law. If we apply Fuller’s rule of law theory to the judiciary as an institution, then we can measure the judiciary against Mureinik’s requirement of ‘a culture of justification’ to determine whether it is, in fact, operating under the rule of law. Our judiciary will be upholding its constitutional imperative if it demonstrates a culture of justification not only in its interaction with broader society, but also in its internal interactions. There needs to be a culture of mutual engagement within the judiciary before a culture of justification and the rule of law can be realised by the judiciary. Is the OCJ designed to play a role in the creation such a culture and promote the rule of law?

VI THE JUDICIARY AS AN INSTITUTION UNDER LAW

The OCJ undertakes to uphold the values of openness and transparency and claims that these values will, inter alia, ensure ‘accountability of the Judicial branch of the State to the people of South Africa [and] public confidence in the Judiciary’.105 It recognises, then, that OCJ and its leadership serve to facilitate the giving of account, and that accountability requires communication and justification. In this part, I argue that a culture of justification is weakened by two factors where the judiciary is concerned: first, the OCJ’s own structures and processes, which are not designed to encourage mutual engagement, and, second, a lack of transparency about the structures and processes of the judiciary. The first issue goes to the need for communication and justification within the judicial institution itself. The second goes to the communication which connects the judiciary to the legislature and broader South African society, allowing the judiciary to render account to its ultimate beneficiaries.


104 L Fuller ‘Positivism and Fidelity to Law – A Reply to Professor Hart’ (1958) 71 Harvard Law Review 630; J Brunnée & S Toope Legitimacy and Legality in International Law: An Interactional Account (2010) 21. When citizens are enabled to reason with norms, they can begin to develop what Fuller called ‘fidelity’ not just to particular norms, but to the system of law itself.

105 OCJ Strategic Plan 2015/16 - 2018/19 (note 6 above) at 10.
A Structure and processes

In part IV B above, the current mechanisms for the OCJ’s extensive administrative control over the judiciary were described in detail. For present purposes, the important aspect of all that information is how little the system attempts to maintain a system of mutual engagement. Thus the section of the Superior Courts Act 10 of 2013 headed ‘Judicial management of judicial functions’, which empowers the Chief Justice to oversee and monitor the administration of the courts, is limited to establishing who the manager is (the Chief Justice) and what the functions are which he or she will be managing. There is no regulation of the processes of management.

Similarly, the notice issued by the Chief Justice in February 2014 sets out a clear hierarchy of responsibility and control for the judiciary and prescribes a set of ‘norms and standards’ for its operation. The accountability structure in the hierarchy is one-way – the lower levels of the hierarchy have to account to the higher levels. There is no regulation of how these functions are to be managed or how the persons assigned to monitor the judicial officers’ fulfilment of their responsibilities are to do so. There is also no mention of how the Chief Justice is to account or to whom.

Similarly, in the ‘Norms and Standards’ section, there is extensive detail about how courts should operate but only brush-stroke consideration of the rights and expectations which the judges themselves may have of the institution and its management.

However, our current Chief Justice suggests that a system of mutual engagement is nonetheless in place, despite the dearth of formal, structural mechanisms for such engagement, and that this is ensured by the leadership of the judiciary. In an interview in 2013, Mogoeng CJ dealt specifically with the critique that he was ‘empire-building’. He pointed out that the ‘Judicial Council’ included the Heads of Courts of the provincial divisions and continued:

I work with a team. And they can overrule me, by the way. This is not my judiciary, this is our judiciary. And you know judges speak their mind. Unless it can be demonstrated that my colleagues, who are Heads of Court, are nothing but puppets; they are thoughtless sycophants who will do whatever I say, then there would be substance in [the critique].

The problem with this approach is that it makes the ability of the OCJ’s governing body truly to reflect the views of the institution as a whole dependant on the character of the particular judges appointed as Heads of Court (and their willingness, in turn, to consult with the judges in their divisions). By contrast, the initial preparatory work for the Langa/Chaskalson vision of the OCJ attempted to design the Judicial Council in a way that makes the individual qualities of the Chief Justice and the Heads of Court less important. The institution must ensure ongoing interaction; it cannot be left to the moral strength and bravery of the individuals who happen to form part of it at any one time. The structure of the institution should therefore include obligatory processes which allow for extensive input from judicial officers and even, where appropriate, require approval from the judges for certain decisions. It should also include clear procedures and criteria for the monitoring of the performance of judicial officers. For the strongest possible culture of justification, it could even provide for the evaluation of the Chief Justice by the judicial officers.

---

106 GN 147 GG 37390 of 28 February 2014 (note 73 above).
107 Ibid at para 4.
108 Rabkin (note 14 above).
109 Hugh Corder, as quoted in Rabkin (note 14 above). See also Chaskalson/Langa report (note 4 above) at para 4.
As set out in the introduction, the Judicial Council proposed by the Chaskalson/Langa report would have exercised oversight over the OCJ, even including non-judicial officers in the oversight process.\(^{110}\) But this Judicial Council has not been created. Despite Mogoeng CJ’s reference to it in his 2013 interview, the term ‘Judicial Council’ has not attained a formal meaning, as it is not mentioned in the 2013 Act. The body to which Mogoeng CJ refers is simply the Heads of Court, which has served as an informal sounding board for the Chief Justice over the years.\(^{111}\) As noted above, the Heads of Court had no formal powers, infrastructure or budget.\(^{112}\) The Superior Courts Act 10 of 2013 does give the body some power: a majority of its members must support any protocols or directives which the Chief Justice issues to the judiciary on matters concerning norms and standards or the ‘dignity, accessibility, effectiveness, efficiency or functioning of the courts’.\(^{113}\) Furthermore, their consent is required for the setting of recess periods.\(^{114}\) Beyond this, the legislation does not accord the Heads of Court a strong role in the day-to-day running of the judiciary as a whole; in particular, the Chief Justice does not need their approval for any requests he makes for funding with respect to the administration and functioning of the Superior Courts.

The proposed model of the Judicial Council currently before the executive describes the Judicial Council as the ‘penultimate authority’ on all issues relating to judicial governance.\(^{115}\) It sets out the Judicial Council’s ‘powers and functions’ as providing policy direction for the OCJ\(^{116}\) and policy guidelines for SAJEI and the Rules Board,\(^{117}\) acting as advisor to the Chief Justice,\(^{118}\) developing norms and standards and evaluating compliance with them,\(^{119}\) and monitoring performance in all of these bodies.\(^{120}\) It will also make representations to Parliament on matters affecting the judiciary.\(^{121}\) While this may present a first step towards mutual engagement, it does not ensure a culture of justification. The most immediate concern is the absence of the magistracy. Instead of including magistrates in the Judicial Council, as suggested by the Chaskalson/Langa report, the latest proposal would relegate them to a secondary advisory body, the Courts Advisory Body, along with the non-judicial officers which the Chaskalson/Langa report initially proposed for inclusion in the Judicial Council itself.\(^{122}\) This is even though the proposal would have the OCJ take over the ‘establishing, monitoring, evaluating and costing norms and standards’ for Magistrates’ Courts,\(^{123}\) and move the court administration of all lower courts from the Department of Justice and Constitutional Development to the OCJ.\(^{124}\)

\(^{110}\) Chaskalson/Langa report (note 4 above) at paras 4.1.1–4.1.2.

\(^{111}\) Ibid at para 3.2.1.6.

\(^{112}\) Ibid at para 3.2.1.6.

\(^{113}\) Superior Courts Act 10 of 2013, s 8(5).

\(^{114}\) Superior Courts Act 10 of 2013, s 9(2).

\(^{115}\) Later Version Committee on Institutional Models Report (note 15 above) at para 5.2.5.

\(^{116}\) Ibid at para 5.2.5.1.

\(^{117}\) Ibid at para 5.2.5.4.

\(^{118}\) Ibid at para 5.2.5.2.

\(^{119}\) Ibid at para 5.2.5.3.

\(^{120}\) Ibid at paras 5.2.5.1–5.2.5.4.

\(^{121}\) Ibid at para 5.2.5.5.

\(^{122}\) Ibid at para 5.3.

\(^{123}\) Ibid at para 4.1.4.3.

\(^{124}\) Ibid at para 4.1.4.5.
Another, important issue bearing directly on the mutual generative engagement that the Judicial Council would need to achieve is that its decision-making mechanisms are opaque, and there is no provision for how often it must meet, and how Heads of Courts should consult the judges within their own divisions. So we are left to deduce how the Judicial Council and OCJ might function from the past and current practices of the OCJ, which we must assume have been guided by the input of the Heads of Court.

Although we have scant information on the internal processes of the OCJ,125 two recent initiatives point to gaps in the processes of communication and consultation within the institution. The first was the ‘Norms and Standards’ document, which, as initially drafted, was met by serious objections, being seen as too prescriptive, detailed and peremptory, and not taking into account ‘the differing needs and circumstances of the various courts’.126 Judges in busy courts were reported to describe the first draft as ‘unworkable’ and an infringement of their autonomy, which, in turn, impacted on judicial independence.127 The second initiative, the case flow management system, was instituted in 2012. This system makes it the duty of judicial officers to ‘intervene as early as possible in the management of a case’.128 Its implementation has been described in the following terms:

A matter allocated to a judge for comprehensive case management is placed entirely under the management of the case management judge who intervenes and assists the parties to resolve interlocutory disputes until the matter is ripe for trial. The case management Judge convenes and presides over as many case management meetings as he or she finds necessary and may give such directives with regard to the matter as the judge finds necessary. The process of case management is completed when the judicial case manager certifies the matter ready for trial.129

The main advantage of this system is that it can speed up the processing of cases.130 Amongst its disadvantages, it increases the work-load of judges and administrative staff, particularly filing clerks, and requires administrative expertise which may not be available.131 In 2015, at the Gauteng Local Division and Eastern Cape Local Division, Mthatha, one judge was allocated to sit every day of the week to preside over pre-trial conferences; in the case of Gauteng, this involved approximately 100 matters.132 Not surprisingly, lack of ‘buy-in’ from judicial officers was identified as a challenge from the inception of the Case Flow Management Project133 and

---

125 See part IV B below.
126 Rabkin (note 14 above).
130 N Manyathi-Jele (note 128 above).
131 The absence of staff is in part due to the low wages offered for these positions. N Manyathi-Jele (note 128 above). See also Draft Report on Case Flow Management Project (note 128 above)(Complaints about lack of storage space and filing facilities); The Report of the KwaZulu Natal Local Division, Durban; Report by the North West Division (On ‘staff constraints’).
has continued to varying degrees over the years.\textsuperscript{134} A draft report to the Case Flow Management Committee of 2017 also reveals the extent to which the project could be set up in such a way that the OCJ intervened directly in divisions of the High Court, without consulting or even informing the judicial officers affected. Thus, at the Gauteng Local Division, the registrar initially appointed to run the project reported directly to the OCJ without involving the leadership of the Division. The judges themselves were caught unawares by the various steps and requirements of the project as it unfolded.\textsuperscript{135}

Both these experiences raise the possibility that the leadership of the OCJ is centralising control of the judicial function without heeding sufficiently the individual working environments and practical obstacles faced by judicial officers. If this is the case, then it is possible that governance as enforcement might not even ensure efficient service delivery, but it clearly affects the autonomy of individual judges and may thus reduce the independence of the institution as a whole.

### B Transparency in process and design

In part VII A, I argued that the rule of law requires both an internal and external culture of justification for the judicial institution. The governance structures of the institution must facilitate the communication of concerns from the judicial officers on the ground to the leadership. But these internal design features need to be made known to the broader public, which is where the OCJ’s respect for openness and transparency must find expression. It may be that the judiciary is, in fact, a well-oiled machine of mutual generative activity, as the depiction by Mogoeng CJ above would suggest. We will only know when the design features of the OCJ are formalised. The judiciary cannot communicate with, and justify itself to, the South Africa public without the transparent institutional structure which ensures that the internal process of communication is taking place. Outsiders cannot see that the judiciary is functioning under the rule of law if they have no idea how it is functioning at all.

The lack of formal institutional design and transparency was challenged over four years ago, when the Secretary-General of the OCJ presented her report to the parliamentary Portfolio Committee on Justice and Correctional Services on 6 May 2013.\textsuperscript{136} At this point, the wording of the Superior Courts Bill had already been settled.\textsuperscript{137} But as one of the committee members, Dene Smuts, pointed out, it did not provide a legislative basis for the administration of the courts.\textsuperscript{138} Instead, it merely entrenched ‘control of the judicial function by the CJ and JPs’.\textsuperscript{139}

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{134} Draft Report on Case Flow Management Project (note 128 above); Report of the Gauteng Local Division; Report of the Eastern Cape Local Division.
\item \textsuperscript{135} Draft Report on Case Flow Management Project (note 128 above); Report of the Gauteng Local Division.
\item \textsuperscript{136} Office of the Chief Justice 2013 Plans: briefing by Secretary General (6 May 2013), available at https://pmg.org.za/committee-meeting/15875/.
\item \textsuperscript{138} Office of the Chief Justice 2013 Plans: Briefing by Secretary General (note 136 above).
\item \textsuperscript{139} Ibid.
\end{itemize}
\end{footnotesize}
For her part, the Secretary-General acknowledged that governance structures still had to be set up, and that the OCJ was researching which model would best suit the judiciary in South Africa. Six years later, we still have no legislation in place to establish how the judiciary is governed. In her 2018 report to Parliament, the current Secretary-General of the OCJ admitted that ‘not much progress’ had been made on the Court Administration model. The OCJ functions behind a veil, with almost no formal regulation of its internal processes. Indeed, uncovering what it is doing and how required some journalistic determination in the case of this article. It is noteworthy that the list of functions which the DOJ transferred to the OCJ in April 2015 was not set out in a Government Gazette, but in a letter from the Minister of Justice to the CJ. The letter is also not available on the website of the OCJ. Similarly, the existence of one particular set of objections to the Norms and Standards was public knowledge, but I was unable to obtain a copy of them despite requests to judges, the Constitutional Court and the OCJ itself.

VII CONCLUDING REMARKS

South African jurists and theorists have recognised that judicial independence is protected by security of tenure, financial independence and administrative independence. There is also an acknowledgement that these three ‘core characteristics’ can support the independence, either of the judiciary as an institution, or of the individual judges. Current jurisprudence and scholarship fail, however, to engage with the relationship between individual and institutional independence, and struggles to identify mechanisms of protection for the institution as such. The factors which have received the most emphasis are the financial independence of the judiciary and the judiciary’s control over its own administration.

The OCJ has not, as yet, managed to secure significant control over the process of determining what funds are allocated to it from central government. This is one of the consequences of its failure to move to phase two of the initial OCJ plan. Where administration is concerned, on the other hand, the OCJ has taken over wide areas of control. However, it is not apparent that the vastly increased control enjoyed by the leadership of the judiciary has translated into improved control for individual judges. We therefore need to determine how the independence of individual judges relates to the independence of the institution.

In this regard, I drew on Lon L Fuller’s account of the rule of law, an approach closely linked to Etienne Mureinik’s concept of a culture of justification. Fuller’s theory suggests conditions precedent for any society or institution, including the judiciary, to operate under law. In particular, it requires what has been termed ‘interactional law’ – a process of mutual engagement – within those institutions which have to uphold the rule of law. Unless the judiciary is itself constituted by law, it will not be equipped to uphold the rule of law in its dealings with litigants, other branches of government and the society which it is meant to serve.

141 I was able to obtain a copy of this letter through the kind help of Yvonne van Niekerk of the OCJ administration. As of 20 January 2019, the letter is still not accessible on the website of the OCJ.
142 One judge did not consider the objections confidential, but did not have a copy. I was then referred to two other judges or former judges who refused to provide the document on the basis that it was internal and confidential. I received no response from the Court or the OCJ when I asked various functionaries at these institutions for the status of the objections, permission to view them and a copy of them.
143 DOJ Policy Framework (note 40 above) at para 71.
If individual judges are not drawn into a process of mutual engagement in the very functioning of the judiciary but instead ‘managed’ by a one-way projection of authority, then the judiciary as an institution is not able to hand down law. Looking at the OCJ from this perspective, it appears that the institution may not be in a position to protect the institutional independence of the judiciary, because it does not contain the mechanisms to accommodate the input of individual judges on the best conditions for effective and independent work. In this respect, I drew on two events within the work of the CJ and the OCJ – the drafting of the Norms and Standards and the Case Flow Management projects – to find indicators of communication gaps within the judiciary. There were insufficient formal sources to describe the current structure and processes of the OCJ, a lacuna which in itself suggests an accountability deficit. The lack of transparency in the functioning of the OCJ may have serious repercussions for judicial independence in South Africa.

Judicial independence is more than the relationship between the judiciary and outside governmental bodies or the general public. It starts, and is maintained by, the rule of law processes within the judicial institution itself. However, its primary virtue is to protect the society which the judiciary serves, which means that the processes of communication, justification and interaction also need to be carried out between the judiciary and broader society on an ongoing basis. This will be possible only when the governance structures of the OCJ are clear and accessible to the South African legislature and public.
The Rights of the State, and the State of Rights in
State Information Technology Agency Soc Limited
v Gijima Holdings (Pty) Limited

ROBERT HLATHI FREEMAN

ABSTRACT: In Gijima, the Constitutional Court held that organs of state cannot review their own administrative decisions by relying on either the right to just administrative action, or the Promotion of Administrative Justice Act (PAJA). This article explores the reasoning behind this finding, suggesting it is sourced in the Court’s understanding of the application of the rights in the Bill of Rights, rather than in a misconstrual of the Constitution’s standing provisions, or an interpretation of s 33 and PAJA. It then considers the Court’s substantive claims about the relationship between the state and private persons, and the nature of the state as a prospective rights-bearer. It concludes by suggesting that thinking of the state as a collection of ‘processes’, rather than as a discrete entity, allows for a more flexible understanding of the application of rights. This flexibility opens space for the state to rely on constitutional rights, even against itself.

KEYWORDS: application of rights, s 8 of the Constitution, rule of law, administrative law, PAJA, judicial review, Bill of Rights, organs of state

AUTHOR: Robert Hlathi Freeman is currently reading for a DPhil at Oxford University, and was previously a researcher at the South African Institute for Advanced Constitutional, Public, Human Rights and International Law (SAIFAC). Email: robhfree@gmail.com

ACKNOWLEDGEMENTS: The author gratefully acknowledges the input of fellow researchers at SAIFAC: David Bilchitz, Mispa Roux, Raisa Cachalia, Justin Wanki and Naomi Hove. Thank you also to Stu Woolman, Salim Nakhjavani and Andrew Konstant for their support, and to two anonymous reviewers, whose contribution greatly improved the piece.
‘I can’t explain myself, I’m afraid, Sir, because I’m not myself you see.’
—Lewis Carroll, *Alice in Wonderland*

I. INTRODUCTION

Are organs of state entitled to claim constitutional rights? This question sits at the heart of the Constitutional Court’s judgment in *Gijima*.¹ Madlanga J and Pretorius J found for a unanimous Court that because organs of state are not protected by the Bill of Rights, they cannot review their own administrative decisions by relying on the right to just administrative action, nor can they rely on the Promotion of Administrative Justice Act (PAJA).² Instead, they ought to pursue review under the rule of law.³ Administrative decision-makers may – and in some cases must⁴ – review their own administrative decisions if wrongly made, but only as breaches of the Constitution’s expectation that all exercises of public power be lawful.⁵

The Court’s finding has garnered criticism because it narrows the domain of rights, while simultaneously expanding the ambit of rule of law review. It is by now customary around law-school water-coolers to remonstrate over the proliferation of the principle of legality, and for good reason: the precise and detailed provisions of PAJA provide a consistency and nuance that rule of law review seldom exhibits.⁶ In a similar vein, some ten years ago in *The Amazing, Vanishing Bill of Rights*, Stu Woolman warned that the sort of rights-avoidance strategy epitomised by *Gijima*, ‘grants the court the licence to decide each case as it pleases, unmoored from its own precedent’.⁷ The increasingly creative strategies for disappearing doctrinal law in favour of discretionary legal principles suggest that the reasons behind the Court’s evasions are in need of address.

At the same time – and unlike in other ‘avoidance’ cases – Madlanga J and Pretorius J do confront the question of whether the Bill of Rights and PAJA should be applied head-on.⁸ The Court is moreover careful to narrow the ambit of its precedent to cases where an administrative decision-maker seeks review of its own decision; organs of state can still administratively review each other’s decisions, and even different decision-makers within the same organ may be entitled to rely on PAJA.

The core contention of this paper is that the foundational problem with the judgment – and the most pressing cause for jurisprudential disquiet – is not the Court’s eschewing of administrative review in favour of the principle of legality, but instead its distorted view of the application of the rights in the Bill of Rights. In short, *Gijima* both misconstrues rights as

---

¹ *State Information Technology Agency v Gijima Holdings* 2018 (2) SA 23 (CC), 2018 (2) BCLR 240 (CC)(*Gijima*).
² Ibid at paras 18 and 37.
³ *Albutt v Centre for the Study of Violence and Reconciliation* 2010 (3) SA 293 (CC), 2010 (5) BCLR 391 (CC) at para 49; *Affordable Medicines Trust v Minister of Health* 2006 (3) SA 247 (CC), 2005 (6) BCLR 529 (CC) at para 49.
⁴ *Merafong City Local Municipality v AngloGold Ashanti Limited* 2017 (2) BCLR 182 (CC), 2017 (2) SA 211 (CC) (*Merafong*) at para 58; *Khumalo v Member of the Executive Council for Education: KwaZulu Natal* 2014 (5) SA 579 (CC), 2014 (3) BCLR 333 (CC)(*Khumalo II*) at para 45.
⁵ *Gijima* (note 1 above) at para 38.
⁸ *Gijima* (note 1 above) at para 30.
constitutively bound up with the relationship between private persons and the state, and adopts
an outmoded conception of the ‘nature’ of the state for the purposes of s 8 of the Constitution.9
These descriptive inaccuracies result in organs of state being barred from relying on the right
to just administrative action, and associatively on PAJA.

The Court’s reasoning gives rise to an application problem because it speaks to who bears
– and who is bound – by constitutional rights. Its key finding is that rights only apply when a
private person claims one of the substantive entitlements in the Bill of Rights. Rights are the
claims of private persons constitutively: what they are is contingent on how they are enforced.
Because of this, the state can never rely on the right to just administrative action, no matter the
breadth of the Constitution’s standing provisions. Addressing the version of rights application
adopted in Gijima is important both because the Court’s perspective impacts administrative
review, and because its conception of the state, and the state’s relationship with private persons,
has implications for determining the ambit of other rights.

Part II of this note sketches the Constitutional Court decisions that led up to Gijima. Part
III considers the reasoning of the Court. Part IV demonstrates that Gijima’s conclusion is
grounded in ideas about the application of the Bill of Rights, rather than on a misconstrual of
standing, or an interpretation of s 33 and PAJA. Parts V and VI then consider Madlanga J and
Pretorius J’s substantive claims about the relationship between the state and private persons,
and the nature of the state as a prospective rights-bearer.

II LEGAL BACKGROUND

The judgment in Gijima ends a dispute that has simmered in South African law for some
time. The Constitutional Court faced it for the first time in Khumalo.10 Mr Khumalo had been
promoted to a post three levels above his extant pay grade in the Kwazulu-Natal Education
Department, seemingly unlawfully. Following a protracted labour dispute brought by
those overlooked for the position, the MEC for Education sought to review Mr Khumalo’s
promotion. The majority of the Constitutional Court found that the ‘true nature’ of the
application was one of ‘judicial review under the principle of legality’.11 By contrast, the
minority decision of Zondo J found that ‘the procedure for bringing [the] application to
Court was governed by the PAJA’.12 The majority itself however finds that should the granting
of the promotion have been administrative action – as opposed to an act in the dominion of
a labour relationship – then the MEC would have had to bring the review through PAJA.13
While steeped in ambiguity, both the majority and the minority decision therefore suggest
that organs of state are, in general, permitted to rely on the right to just administrative action.

A similar position obliquely emerges in Kirland Investments decided three months after
Khumalo.14 In that case the Superintendent-General for Health in the Eastern Cape (Mr Boya)
was confronted by a decision, made in his absence, licensing the establishment of a private

---

9 Section 8(4) of the Constitution of the Republic of South Africa, 1996 (Constitution) states: ‘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.’

10 Khumalo II (note 4 above).

11 Ibid at para 28.

12 Ibid at para 92.

13 Ibid at para 27.

14 MEC for Health, Eastern Cape v Kirland Investments 2014 (3) SA 481 (CC), 2014 (5) BCLR 547 (CC)(‘Kirland’).
hospital. The decision was taken, apparently unlawfully, by an acting Superintendent-General who had been appointed while Mr Boya was on leave. The Court was asked to determine Mr Boya’s responsibilities when faced with the alleged unlawful act on his return to work. In deciding that Mr Boya had a duty to review the decision – as opposed to simply ignoring it – the majority of the Court relied on a close reading of several of PAJA’s provisions. It also grounded its reasoning in the constitutional obligation on the state to ‘provide for review of administrative action’. This suggests that the court saw Mr Boya’s review of the acting Superintendent-General’s decision as the vindication of the right to just administrative action through its statutory extension: PAJA.

The opacity inherent in Kirland and Khumalo became more acute in a triptych of cases, decided within a year of each other. The first of these, Merafong, involved a decision of the Minister of Water Affairs and Forestry to overturn a previous decision made by the Merafong Municipality placing a surcharge on the price that AngloGold paid for water in the District. Deciding that the Municipality had been obliged to obey the Ministers decision – lawful or otherwise – the majority of the Court reasoned that PAJA contained provisions that required the Municipality to institute review even if the administrative decision was patently unlawful. As with Kirland, this line of argument suggests that a state entity – exemplified here by the Municipality – should bring a review in terms of PAJA when faced with an unlawful decision by an organ of state.

The majority made clear, however, that it was not deciding the question of whether PAJA or ‘legality’ was the appropriate avenue of review. When considering Merafong’s tardiness in challenging the Minister’s decision, Cameron J noted that ‘Whether under PAJA, or legality review, [Merafong] was obliged to institute proceedings to review the decision without unreasonable delay.’ This suggests that the Court had no intention of prescribing the pathway to review that Merafong should have utilised. Later, the majority emphasised that the application the Court was asked to decide had been brought by AngloGold in order to enforce the Minister’s decision, rather than being a review on behalf of Merafong. For this reason, the Court did not need to determine if PAJA applied.

Faced with a similar legal quandary in Tsimba, the Court once more sidestepped the question of which pathway to review was apposite. In that case, the Director-General of the Department of Transport extended a contract with a private company, Tsimba, to provide road management services in Gauteng. The extension was, again, granted on tremulous legal grounds. Under duress from Tsimba to perform in terms of the contract, the department sought to subsequently reactively challenge its own decision. Unlike in Merafong, the Court considered, and upheld, the Department of Transport’s reactive challenge. It did not, however, decide under what auspices the department’s review should have been brought. In a footnote, the Court instead found that: “The question of whether challenges, reactive or otherwise, to exercises of public power by the state can be initiated through PAJA need not be decided here.”

15 Ibid at para 82.
16 Ibid.
17 Merafong (note 4 above).
18 Ibid at para 73.
19 Ibid at para 75.
20 Department of Transport v Tsimba 2017 (2) SA 622 (CC), 2017 (1) BCLR 1 (CC)(Tsimba).
21 Ibid at fn 78.
Unlike Merafong and Tasima, the third of this triptych, Aurecon South Africa, did take the form of a direct review of its own decision by an organ of state.\footnote{City of Cape Town v Aurecon South Africa 2017 (4) SA 223 (CC), 2017 (6) BCLR 730 (CC)(‘Aurecon’).} Here, the City of Cape Town sought to review its own decision to award a contract to decommission a power station to Aurecon. While opting to use the 180-day rule set out in PAJA as a benchmark for determining if the review had been brought too late, the Court made explicit that it was not determining the question of whether PAJA or the rule of law was the preferable pathway to review chiefly because arguments on the question were not properly before it.\footnote{Ibid at paras 34–37.}

These cases suggest that while no definite pronouncement had been made, the Constitutional Court more often than otherwise has assumed that PAJA could be used where the state sought to review its own administrative action. Remarkably, in none of the cases was the question of whether or not the state is entitled to rely on rights as a whole even canvassed. Instead, emphasis was always put squarely on whether or not the decision itself was administrative action, and whether any sort of review could be brought. It was therefore something of a surprise when the Gijima Court reversed the prevailing view.

### III Gijima, The State and Rights

Gijima, a private company, contracted with the State Information Technology Agency (SITA), an organ of state, to provide IT support to the South African Police Services (SAPS). Primed by the imminent threat of litigation to enforce the contract, SITA sought to review the tender and set it aside on the grounds that proper procurement mechanisms had not been followed when it was awarded.

The High Court found that the granting of the tender constituted administrative action under PAJA, and that the review should consequently be determined in terms of that Act. The Supreme Court of Appeal agreed. It also confirmed that the principle of subsidiarity required that the review be conducted in terms of PAJA, rather than directly through s 33 of the Constitution, or under the rule of law.

Madlanga J and Pretorius J, writing for a unanimous Court, came to a different conclusion. They found that an organ of state can never rely on PAJA, nor the right to just administrative action, in seeking to review its own decision. Instead, SITA’s review could only have been brought as a breach of the principle of legality.

The Gijima Court supported its finding by characterising PAJA as an extension of the constitutional right to just administrative action.\footnote{The Court’s claim that, ‘the concept of “administrative action” in whatever section of PAJA cannot suddenly have a meaning wider than that envisaged by the source of the concept, namely the Constitution’, is troubling in its own right: legislation – even constitutionally mandated legislation – can surely offer more entitlements than the Constitution sets out.} It then reasoned that as s 33 is a right, the Court must first determine the ‘philosophical underpinnings of the very notion of whom fundamental rights are meant to protect’.\footnote{Gijima (note 1 above) at para 18.} As the Court saw it, whether SITA is entitled to rely in general on the rights in the Bill of Rights should ‘inform’ the interpretation of the right to just administrative action.\footnote{Ibid.} With this in mind, Madlanga J and Pretorius J find it ‘quite axiomatic that fundamental rights are meant to protect warm-bodied human beings primarily...
against the State. They who are not ‘human beings’ (seemingly inclusive of private juristic persons), are therefore not entitled in general to directly invoke rights. As organs of state like SITA fall into this category, they cannot have recourse to the right to administrative justice, and so cannot rely on PAJA when orchestrating review.

The Court substantiated its conclusion in two ways. First, through an analysis of existing decisions and the context in which the rights-regime in South Africa has emerged, it looked to show that the Constitution only ever meant to afford private persons rights, essentially for purposes of responding to state action or inaction. They began by citing the First Certification judgment, in which the Court found that, ‘The movement to recognise and protect the fundamental rights of all human beings gained increased momentum in the international arena from the end of the Second World War.’ The Court put emphasis on the words ‘human beings’ to find that, from the outset, South Africa’s constitutional rights were aimed at protecting individual persons and their collective associations. They noted further that the Constitutional Principles that gave rise to the Bill of Rights in the Interim Constitution were described by the First Certification Court ‘as the inalienable rights of human beings’ (Italics in the original).

Thereafter the Court genealogically connected the rights in the Bill of Rights, through the Constitutional Principles, to the ‘fundamental rights and freedoms’ enshrined in the United Nations Universal Declaration of Human Rights (UDHR) and other international human rights instruments. Madlanga J and Pretorius J construed these international entitlements as ‘universal’ only to the extent that all human beings have them against the state. Noting that South Africa’s Bill of Rights protects different rights to those protected by international legal instruments, Madlanga J and Pretorius J nevertheless found that the rights in the Bill of Rights maintain the same character of applying only to private persons.

Second, the judgment asserts that it is ‘inconsonant that the State can be both the beneficiary of the rights and the bearer of the corresponding obligation that is intended to give effect to the rights.’ In other words, as a matter of principle, the state cannot be both the giver and receiver of the same legal entitlement. Madlanga J and Pretorius J argued that this is especially true in the case of the right to just administrative action, where to argue otherwise would mean that an organ of state would be entitled to ask itself for reasons, or that it act reasonably, or in accordance with the law.

Finally, the Court turned to interpreting the content of s 33 and PAJA against the background of its previous analysis. Madlanga J and Pretorius J tracked the abusive history of administrative law in South Africa. Noting that the object of this abuse had been private persons, and the subject the state, they found that ‘the intended beneficiaries of the change’ in South Africa’s approach to holding administrative decision-makers to account ‘were private
persons’.35 For this reason, they concluded that the right to administrative justice existed solely to protect private persons from the abuse of state-sanctioned administrative decision-makers. Resultantly, organs of state cannot utilise PAJA.

IV THE APPLICATION OF RIGHTS

Several perceptive commentators have read Gijima as making a finding about whether or not the state has standing to enforce PAJA. Both Mitchell de Beer and Leo Boonzaier show that neither the provisions relating to standing in the Act, nor s 38 of the Constitution, preclude the enforcement of a PAJA-right by a person different from the right-holder.36 It seems clear that if Gijima aims to ‘establish that organs of state lack standing to enforce PAJA’s rights’, then it should fall foul of these challenges. The standing of organs of state should not depend on whether the state’s interest is at stake, and, even if it were, there is little reason to think that SITA lacks sufficient personal interest in the relief sort.

My view, however, is that this reproach is at odds with the conceptual shape of the judgment. As the previous section outlines, the bulk of the Court’s analysis has to do with the historical and philosophical roots of rights in general. The Court deliberately began its reasoning by determining the nature of rights as a whole. Madlanga J and Pretorius J subsequently used this framework to argue for rights being inherently privileges claimed by private persons, ‘primarily’ from the state. In exceptional cases, they are claims made by private persons against one another. But, pivotally, rights are always claims made by private persons of some or other party. The substantive entitlements in Chapter 2 of the Constitution only become ‘rights’ when private persons do the claiming. The Court spent considerable energy establishing this point, and fleshing out its historical genesis. It saw the notion as both an ‘axiomatic’ truth, and a feature of existing law, italicising ‘human beings’ at every opportunity.37 It was also the norm it hoped to extract from the First Certification case, and what it characterised as a ‘fundamental right’ in Constitutional Principle II. Madlanga J and Pretorius J’s contention was that the question of whether the state can claim rights is dependent, firstly, on when rights are claimable, and only subsequently on who has standing to enforce them. The animating structural idea is that rights are not free-floating entitlements; they are relational: rights lack conceptual force without reference to the relationship between private persons and either the state or other private persons. Reference to the question of who rights ‘protect’ is a misnomer – the substance of the Court’s point is about when rights can be claimed at all.

This reasoning evokes the language of standing. Who can enforce the substantive entitlements in the Bill of Rights remains central to the Court’s reasoning. But what is distinctive about its finding is that enforcement questions are built into its understanding of rights. More accurately, the application of the claims in the Bill of Rights is defined in terms that include reference to their enforcement. As I see it, the Court finds that rights only apply when private persons are doing the enforcing. Madlanga J and Pretorius J began their

37 Ibid at para 18.
discussion of the ‘philosophical underpinnings of the very notion of whom fundamental rights are meant to protect’ by referring to the assessment of s 8 of the Constitution in the First Certification judgment, and application remains at the heart of their analysis. The state cannot enforce rights claims not because it lacks standing, but because rights only apply when the substantive entitlements in question are claimed by private persons: who claims the right is a constitutive feature of when rights apply.

That this is the line of reasoning is not always plain. Gijima actively fudges findings about interpretation with findings about application, which obscures its insouciance to standing still further. When the Court shifts from its analysis of the meaning of ‘fundamental rights’ – an inquiry essentially about application – to its construal of s 33 – a finding about interpretation – it claims to be moving from sketching the ‘background’ of the Bill of Rights into defining the ambit of the right to administrative justice. But if this background incorporates claims about when rights apply in general – as I have suggested is the case – then the Court is eliding application and interpretation. Application sets the scope and form of the Bill of Rights. Interpretation establishes the meaning of the rights themselves. The interpretation of a particular right does not define the application of rights as a whole; the application of rights instead sets the parameters of that right’s interpretation. Shoehorning an application finding into an interpretation analysis allows the Court to avoid taking on the full implications of its decision for the ambit of rights in general, while simultaneously acquiring fuzzy reasons for its specific finding that the state is barred from relying on PAJA as an ‘extension’ of s 33.

This is not to say that application and interpretation should be hermetically sealed from each other. Section 8(4) of the Constitution states that a juristic person is entitled to the rights in the Bill of Rights ‘to the extent required by the nature of the rights’. Courts are therefore required to interpret individual entitlements in order to establish their application, and there may be occasions when the application of a right is dependent on its content. This is because there are some rights – for example the right to life or dignity – which a juristic person cannot sensibly hold. These are cases where the general application of the Bill of Rights bends to the specificities of particular entitlements. But this is different from saying that the interpretation of a specific right can impact on how all rights apply. Just because PAJA can be interpreted to operate between private persons and the state, does not mean that there is a general rule of application to this effect.

There are also strong reasons for suggesting that s 33 cannot be interpreted to exclude the state from relying on the right.38 Various constitutional provisions explicitly limit the ambit of rights to ‘citizens’, ‘workers’, ‘children’, and so on. Using interpretative tools to determine what these terms include, and consequently who is entitled to the associated privileges, is a regular feature of constitutional adjudication. In Gijima, the Court purported to be doing just this sort of interpretative work when it found that ‘everyone’ in s 33 of the Constitution does not include organs of state.39 It argued that as s 33(3) ‘imposes a duty on the state to give effect to the rights in s 33(1) and (2)’ (Court’s italics), it would be ‘inconsonant’ for the state to also be the beneficiary of the right.40

39 Gijima (note 1 above) at para 27.
40 Ibid.
Notably, however, the facts of Kirland, Merafong and Tasima illustrate that, in practice, there are many occasions when the state does have reasons for wanting to review its own decisions; while Aurecon, as well as Gijima itself, show that this ambition can even extend to the same decision-maker. These cases often arise in the context of mismanagement or corruption, and the review is generally brought to avoid the consequences of a past wrong. The important point is that the Court has entertained the review (or rejected it on the basis of delay) in each case. The Gijima Court was also happy to accept the principle, first established in Kirland, that organs of state must, on most occasions, frontally challenge their own decisions. It was therefore not opposed to organs of state reviewing their own decisions as a general proposition.

Of course, if the entitlements in PAJA – an Act of Parliament – can be (and are) wider than s 33, the Court’s point about the ambit of the rights in the Bill of Rights is largely academic. What matters is whether SITA’s awarding of the contract constituted administrative action as defined by PAJA, rather than who it is that makes the claim. Whether or not the rights in the Bill of Rights are defined by their enforcement by private persons has seemingly little to do with whether Parliament can give the same substantive powers to the state, without limiting its enforcement. De Beer makes the point in helpfully rhetorical form: ‘had Parliament explicitly required organs of state to rely on PAJA when seeking to have their own decisions set aside on review, would that be unconstitutional? I would argue “no”.’41 This argument is essentially convincing. PAJA itself makes unnecessary the Court’s findings on the application of rights: its reasoning should have started and ended with the Act. It does, however, raise the spectre of the relationship between constitutionally obligatory legislation and the constitutional provisions – as well as broader social values – that they cement.42 Lael Weitz argues compellingly that constitutionally obligatory legislation has special status, because it affords the legislature the power to augment the Constitution. But can a piece of legislation, explicitly promulgated to define and protect a right, take on a shape fundamentally different to that of the corpus of rights in general? Put otherwise, if Madlanga J and Pretorius J are correct that rights only apply between private persons and the state, can PAJA broaden this scope to include relationships between private persons and the state?

These are delicate questions about how constitutionally obligatory legislation affects the way in which the Constitution is interpreted and applied. Whether constitutionally obligatory legislation should be applied and interpreted as ordinary legislation, or as a constitutional provision, or some hybrid of both, has not been settled in our law. AgriSA suggests that individual constitutional rights should be interpreted in ways that are compatible with legislatively-determined content where that legislation is constitutionally mandated.43 Extending this maxim to the application of rights as a whole nevertheless seems a substantial leap, handing considerable power usually set aside for the judiciary to Parliament.

In any event, the impetus behind Gijima – and the reason that the Court ultimately concludes that PAJA does not apply – is the finding that a right is a claim made by private

---

41 De Beer (note 36 above) at page 620–621. Boozaier (note 36 above) at page 650 makes a similar point: ‘[T]o insist that the legislature may not enhance those protections is a different thing altogether. That would be a self-evidently unsound principle, thoughtlessly curtailing the legislature’s rightful use of its powers.’

42 Section 33(3) of the Constitution requires that national legislation be enacted to give effect to the administrative law rights set out in section 33(1) and 33(2). As the preamble explains, PAJA was promulgated for this purpose: ‘To give effect to the right to administrative action that is lawful, reasonable and procedurally fair and to the right to written reasons for administrative action as contemplated in section 33 of the Constitution.’

persons. In the next section, I suggest that this approach is nevertheless dependent on an understanding of the state and its role in defining rights. The Court’s reasons for wanting to construe rights as defined as the claims of private persons is premised on an perception/view of the state as an homogeneous purveyor of rights. I conclude in the final section by suggesting that a better approach may be to consider the question of whether a state-actor can rely on the right to just administrative action in – to use an abused phrase – on a ‘case-by-case basis’, with emphasis put on the process under review, rather than the institutional actor.

V RIGHTS AS THE CLAIM OF PRIVATE PERSONS AGAINST THE STATE

On its face, the Court’s finding in Gijima that the nature of rights is such that they only apply when claimed by private persons can be excised from any dispute over whether or not rights are bound up with the relationship between the state and private persons. It is possible that Madlanga J and Pretorius J are correct that rights are the entitlements of private persons alone, but wrong that rights exist primarily between private persons and the state. The Gijima Court seems, however, to interweave the two arguments. It explains that the reason why the state cannot rely on rights is because rights exist to off-set the threat posed by the state to private persons. On the Gijima reading, the whole schema of rights is dependent on holding state power to account — if there was no political force exerted by the state on private persons, there would be no need for rights.

To further engage this reasoning, the following section looks at the conception of the state in the application of rights. Gijima’s historical analysis rightly shows that viewing rights as the claims of private persons against the state has deep historical roots: the discourse of human rights in its modern form emerged as a response to state oppression. But the record of rights in the years since the passing of the UDHR has brought this rigid structuring into question. In particular, proponents of rights now generally acknowledge that the power dynamics policed by rights do not exist solely between the state and private persons. Indeed, with the rise of deregulated economies, it is increasingly the relationships between private persons themselves that need supervision.

A prescient example is the rise of international agreements on business and human rights. Article 12 of the Guiding Principles on Business and Human Rights is headed: ‘The responsibility of business enterprises to respect human rights’. This responsibility includes businesses avoiding, causing or contributing to human rights abuses, and acting to mitigate the negative human rights consequences of their work. South Africa has committed itself to

---

44 Gijima (note 1 above) at paras 18 and 26.
45 An important facet of the Gijima court’s approach is that it necessarily associates rights and institutions. There may be moral or agency-based support for what a private person is entitled to ask of the state, but these are not free-floating entitlements. They exist because people and the state exist. This is a political conception of human rights: we identify what rights are by illustrating what role they play in some political sphere. C Beitz The Idea of Human Rights (2009); J Rawls The Law of Peoples (1999).
46 Gijima (note 1 above) at paras 26–31.
49 Guiding Principles on Business and Human Rights, art 12.
50 Ibid art 13.
developing a binding international treaty on Business and Human Rights, suggesting awareness on the part of the Executive that human rights obligations extend beyond the relationship between the state and private persons.

More tellingly, that the application of the Bill of Rights extends beyond the vertical private-state dichotomy is embedded in s 8 of the Constitution. Section 8(2) states that, ‘A provision of the Bill of Rights binds a natural or a juristic person if, and to the extent that, it is applicable, taking into account the nature of the right and the nature of any duty imposed by the right’. In *Khumalo v Holomisa*, the Court found this to mean that at least some rights in the Bill of Rights bind private persons some of the time.\(^{51}\) In *Juma Musjid Primary School v Essay*, the Court extended this interpretation to ‘require private parties not to interfere with or diminish the enjoyment of a right’.\(^{52}\) Recently, in *Daniels v Scribante*, this finding was extended further to mean that, ‘the duty imposed by the right to security of tenure, in both the negative and positive form, does rest on private persons’.\(^{53}\) These cases show that there are at least some occasions where the Constitution requires both the positive and negative horizontal application of rights – it is simply not true that rights are inherently claims made by private persons against the state.

There are also strong theoretical reasons for holding that rights apply beyond the vertical private-state relationship. Stu Woolman argues convincingly that s8(1)’s use of the words ‘all law’ means that every extant legal artefact is bound by the Bill of Rights.\(^{54}\) This includes common law agreements between private parties, and, on what Woolman calls a ‘Hohfeldian’ view of the law, also the body of norms that govern social relationships outside of explicitly codified rules.\(^{55}\) If this is the case, then it is clear that the Bill of Rights finds application even where the state is not a duty-bearer.

The *Gijima* Court acknowledges this line of work, but does not see it as terminal to its interpretation of the application of rights.\(^{56}\) This reticence invites questions over why, exactly, South African courts continue to want to view rights as inherently caught up with the rapport between the state and private persons. A charitable interpretation of the phrase ‘meant to protect warm-bodied human beings, primarily against the state’ may provide some traction in answering this question. Perhaps the Court means to say no more than that rights inherently provide aegis to private individuals, and that both the most prescient threat to their rights, and the most capable source of their vindication, is the state. This severs the conceptual tie between private persons and the state in defining rights, and erects in its place a looser association sourced in the essential role the state plays in governing a constitutional democracy.

The reasoning in *Blue Moonlight Properties* is an instructive example here.\(^{57}\) In that case, a private company, Blue Moonlight, sort to evict a number of unlawful occupiers from a building it owned. The occupiers objected to the eviction order. After noting that the

---


\(^{52}\) *Governing Body of the Juma Musjid Primary School v Essay N.O.* 2011 (8) BCLR 761 (CC) at para 58.

\(^{53}\) *Daniels v Scribante* 2017 (4) SA 341 (CC), 2017 (8) BCLR 949 (CC) at para 49. See also *Barkhuizen v Napier* 2007 (5) SA 323 (CC), 2007 (7) BCLR 691 (CC) at para 57.


\(^{55}\) Ibid at Ch 31 46–47.

\(^{56}\) *Gijima* (note 1 above) at fn 18.

\(^{57}\) *City of Johannesburg Metropolitan Municipality v Blue Moonlight Properties* 39 2012 (2) BCLR 150 (CC), 2012 (2) SA 104 (CC) (*Blue Moonlight*).
Occupiers grounded their claim in the rights to housing and equality, the Court immediately shifted to determining the state’s responsibility in alleviating the detrimental consequences of the proposed eviction. This approach seems to have been prompted by Blue Moonlight’s protestation that the occupiers’ continued occupation violated the company’s right to property.

The case thus pitted two compelling, and opposing, rights claims against each other. What is more, two opposing rights claims with considerable political heft. In such a context, it is understandable why apportioning responsibility to the state looms as a compelling middle way. By finding that the state had to provide adequate alternative accommodation to the occupiers (on the basis of its own existing laws and policy), the Court was able to shift the rights obligation of housing the applicants from the owner of the building to the state. At the same time, by sanctioning the eviction on condition that the state make alternative accommodation available, the Court similarly shifted the rights obligation of avoiding arbitrary deprivation of property from the occupiers to the state.

The implicit reasoning adopted by the Court is that the threat to the housing rights of the Occupiers, and the property rights of the owners, was a result of the state’s failure to provide adequate alternative accommodation. The Court conspicuously declined to declare that either the occupiers or the owners infringed rights: it was only the state’s housing policy that was found unconstitutional. But does this mean that by bringing the evicting application Blue Moonlight was not violating the occupiers’ rights? And by illegally settling the building, were not the occupiers violating the rights of Blue Moonlight?

Through finding, in this and other cases, that the state must always be joined in eviction applications, the Court has ensured that we will never know the answer to these questions. Instead, the state remains the only visible duty-bearer: the Court allocates duties to the state in order to avoid the more complex question of resolving mutually valid competing rights claims between private persons. State apparatuses become a collectively-agreed – and judicially-sanctioned – mechanism for characterising and resolving social ills.

The judicial signal is that — in cases like Blue Moonlight — there are reasons why private disagreements may be better resolved by allocating the obligations that arise from being bound by rights to organs of state. Doing so enables courts to acknowledge the political and social character of rights disputes while providing solutions that avoid placing responsibility at the door of either of the private litigants.

In many respects, this is a positive feature of our constitutional order. The state is given the obligation to protect and promote rights, even where the claim jeopardising a person’s interests are seemingly private in nature. At the risk of oversimplification, the state collects taxes, and holds the power of coercion exactly so as to bring about the social goods set out by law-makers and the Constitution. The holding of this obligation helps ensure that disagreements and social tensions are diverted through legal channels, rather than being settled outside of them.

There are therefore good reasons for having the state at the centre of the realisation of rights.

Does this analysis mean that organs of state should be barred from relying on rights themselves? My tentative view is that it does not. Indeed, the reverse seems to be true: if the state bears the burden of ensuring the protection of rights, it should be able to apply them directly when doing so is advantageous for ‘human beings’. This is exactly what occurred in

---

58 Ibid at paras 19–20.
Gijima. SITA’s review was grounded in its belief that the contract signed between itself and Gijima did not comply with s 217 of the Constitution, and was therefore unlawful. One of the key normative principles at the heart of s 33 is that unlawful administrative actions should be set aside. The setting aside of unlawful decisions is valuable because it protects private persons from capricious state action. The benefit derived from this standard surely does not change when it is the state that seeks to enforce the law.

The converse view, which sees administrative review as the preserve of private persons, suggests that what matters is not so much the content of the right, as the fact of a private person enforcing it. This interpretation is echoed by the Gijima Court’s finding that it would not make conceptual sense for an organ of state to review its own decisions. Such a position has distinctly political undertones. Rights are here construed as the entitlements of a disgruntled citizenry, spurred on by an antiquated form of republicanism in which the state is pictured as a monolithic iniquity which private persons must hold to account.

Aside from the obvious problems with this fanciful depiction of the work done by a modern state, there is nothing in Gijima which suggests a clear understanding of what it would mean for the state to hold itself to account. In the context of debates on state capture and developing modes of accountability discussed elsewhere in this collection, what the state is, and whether it can claim rights against itself, emerges as an increasingly important concern.

VI CAN THE STATE CLAIM RIGHTS AGAINST ITSELF?

A natural starting point in answering this question is s 8(4) of the Constitution. It provides that juristic persons are 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’. The scope of this qualifier has not yet received attention from the Court. In fact, s 8(4) has largely been ignored by South African courts. Nor does the Constitution itself provide much further guidance. Unlike in the case of determining who is bound by the Bill of Rights, there is no helpful coda – played in the former instance by s 8(3) – which explains how the provision should be applied.

What is notable is the slight difference in wording between s 8(2) and s 8(4). Where the former states that juristic persons are bound by the Bill of Rights to the extent required ‘by the nature of the right and the nature of the duty imposed by the right’, the latter provides that juristic persons are 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’. The difference is that, in the case of s 8(2), emphasis is put on the duty imposed by the right, while in the case of s 8(4), emphasis is put on the nature of the juristic person.

This may suggest a general reluctance to extend the application of the Bill of Rights beyond the relationship between ‘qualifying’ juristic persons – those who exhibit certain (at present unclear) features – and those persons covered by ss 8(1) and 8(2) of the Constitution. But the shift away from the vertical view of rights discussed in the section above remains: anyone can now be both the holder of rights, and the bearer of corresponding duties, if the reasons are good. As the Court puts it in Association of Mineworkers and Construction Union, the Constitution ‘casts up no impenetrable wall between the public and the private’.61 Our legal order rebuffs a doctrinal approach to the application of the Bill of Rights.

61 Association of Mineworkers and Construction Union v Chamber of Mines of South Africa 2017 (3) SA 242 (CC), 2017 (6) BCLR 700 (CC) (‘AMCU’) at para 68.
Assuming – correctly, I believe – that organs of state are juristic persons, what, then, is the ‘nature’ of the state for the purposes of s 8(4)? Specifically, what is the nature of the state at the point it seeks to rely on a right against itself? The Constitution’s definition of organ of state is helpful in this regard. It defines an organ of state as either, ‘any department of state or administration in the national, provincial or local sphere of government’, or any ‘functionary or institution… exercising a power or performing a function in terms of the Constitution or a provincial constitution or exercising a public power or performing a public function in terms of any legislation’.62

The Constitution therefore defines the state either by its institutional position or by its function. This raises the intriguing proposition that an organ of state in the first sense may perform acts that are not public in nature.63 If this is the case, then why should an organ of state not invoke the right to just administrative action in a costume different to that it wore in making the original decision it now seeks to review? In language more faithful to s 8, why should the nature of an organ of state not be different at the point it makes a decision and at the point it reviews it? This is even more credible in the case of an outwardly private party who previously exercised a public power, and now seeks to review its own decision in a private capacity.

The point is that when determining the ‘nature’ of an organ of state for the purposes of s 8(4), the emphasis should be on the function of the decision, rather than the functionary itself. This suggests we may also need to move our understanding of the state away from a collection of definable institutions and entities. This important work of ‘reimagining’ our accepted depiction of the state deserves book length study beyond the ambit of this short diagnostic note. I would like to end, however, with a suggestion of one direction in which this work may go. In my view, rather than seeing the state as an homogeneous whole – a definable thing – s 8(4) challenges us to approach it as an assortment of processes. This would imply that the state is the instantiation of a variety of different social practices. As Nicos Poulantzas explains, the manifestations of the state are the effect of various social productions and contestations, rather than things in themselves.64

Gaston Bachelard provides a helpful metaphor for this reading. He explains that ‘the idea of the state’ is analogous to the pre-scientific idea of fire: flames appear to be an object or a thing – they have ‘palpable confirmations’ – giving rise to substantialist ideas of fire as some essential quality.65 In reality, however, fire is no more than the outcome of physical processes. An account of fire must therefore be in terms that invoke these processes rather than the effects that follow.

Similarly, the state is not an entity in itself, but instead an assortment of varying processes that bring about decisions and visible applications. Determining whether or not an organ of state should be entitled to rely on a right should consequently focus on the underlying processes that manifest themselves in the organ seeking to rely on the substantive entitlements of the ‘right’. This analysis can have fluctuating outcomes – it is a sociological and factual determination rather than a strict rule. Understood in this way, it is perfectly coherent for the

---

62 Constitution s 239.
63 Cape Metropolitan Council v Metro Inspection Services (Western Cape) CC [2001] ZASCA 56, 2001 (3) SA 1013 (SCA); AMCU (note 61 above) at para 69.
64 N Poulantzas State, Power, Socialism (2000).
65 G Bachelard The Poetics of Space (2014).
state-as-process to rely on rights against the state-as-process: this interaction does no more than form part of the set of mechanisms by which the state proceeds.

Such an interpretation is in contrast to the Gijima Court’s view of the state as itself the source of social harms, and therefore the target of the historical changes in the character of administrative law aimed at preventing the abuse of state power.66 The ‘flames and heat’ of the Apartheid regime’s performance of administration results in the court misconstruing the ‘wood’ of a dominant (and nefarious) ideological process for the ‘trees’ of the state. Instead of seeing administrative justice as a mechanism to prevent the abuse of public power, the court sees it as a tool to prevent the abuse of private citizens by the state.

This leaves the Court’s argument that it is ‘inconsonant that the State can be both the beneficiary of the rights and the bearer of the corresponding obligation that is intended to give effect to the rights.’67 Again, the problem only arises if the state is construed as a monolithic entity. As soon as SITA’s decision to review its own decision, and the making of the decision itself, are seen as two separate processes, the difficulty falls away. While the decision-maker may be both the bearer of the right and of its corresponding duty, the processes that lead up to each decision are distinct. SITA as bearer-of-rights is not a mimetic representation of SITA as bearer-of duties.

This is not to say that organs of state should always be entitled to rely on rights. Instead, the limitation of the ambit of the application of rights to state entities should be a dependent fundamentally on the nature of the process (together with the nature of the right), rather than merely on the basis of the state being the state. Judgments like Gijima emerge from a dogmatic application of the reasoned law illustrated by Blue Moonlight. Who should be held to account, and by whom, is a processive inquiry. As the facts of cases like Kirland, Tasima and Aurecon show, the state is made up of competing interests, often inhabiting the same institutional position at different times. The emphasis should always be on how best to uphold the substantive obligations that rights protect. Taking this approach allows rights to function as normative guidelines, and liberates state entities from the strictures of too vertical a vision of law and society.

66 Gijima (note 1 above) at para 26.
67 Ibid.
Breaking the Silence: The Treatment of *Ius Cogens* in Zimbabwe Torture Docket and Al Bashir

**MIA SWART**

**ABSTRACT:** The South African trajectory of the litigation in the *Al Bashir* case ended at the Supreme Court of Appeal (SCA). The fact that the case never reached the Constitutional Court (CC) has significant implications. It means that the apex court did not have the opportunity to make authoritative pronouncements on South Africa’s obligations under the Rome Statute of the International Criminal Court, and that the it did not resolve the question whether sitting heads of state charged with *ius cogens* crimes – that is, crimes prohibited under peremptory norms of public international law – are protected by immunity before national courts. Had the case proceeded to the Constitutional Court and had it highlighted the *ius cogens* nature of the crimes with which Al Bashir is charged, this would have set a correct and important precedent for future cases of this kind. This article examines the flaws in the reasoning of the SCA in the *Al Bashir* case, focusing on the SCA’s thin and inadequate treatment of the relevant norm of *ius cogens*. The recourse to and application of *ius cogens* norms has been underexplored in the South African context. Before the CC, the concept has been neglected, despite the imperative of section 232 of the Constitution. To illustrate this neglect, the cases of *Azapo*, *Wouter Basson* and the so-called *Zimbabwe Torture Docket* case will be analysed. These three cases all concerned the adjudication of international crimes. The *Azapo* and *Wouter Basson* cases can further be described as ‘transitional justice cases’ since they dealt with the prosecution of state-sponsored crime during apartheid. The nexus of these cases with *ius cogens* norms is inescapable; and the silence of the CC especially stark.

**KEYWORDS:** *ius cogens*, norm conflict, immunity, Al Bashir, transitional justice

**AUTHOR:** Research Director, Democracy and Governance Unit, Human Sciences Research Council (HSRC). Email: miaswart26@yahoo.com
The acceptance therefore of the _jus cogens_ nature of the prohibition of torture entails that a state allegedly violating it cannot invoke hierarchically lower rules (in this case, those on state immunity) to avoid the consequences of the illegality of its actions. — _Al-Adsani v United Kingdom_¹

## I INTRODUCTION

When former Sudanese President Omar Al Bashir set foot on South African soil in June 2015, few could predict the legal controversy his presence in South Africa would trigger. Al Bashir, charged by the International Criminal Court (ICC) for war crimes and crimes against humanity in 2009² and for genocide in 2010³, was President of Sudan at the time and no stranger to controversy. Whereas his visits on the continent⁴ and beyond⁵ have stirred debate and a substantial body of jurisprudence⁶, his visit to South Africa prompted litigation starting in the High Court,⁷ proceeding to the Supreme Court of Appeal⁸ (SCA) and (almost) ending in the Constitutional Court. Ultimately, the ICC also pronounced on South Africa’s non-compliance with its obligations under the Rome Statute.⁹

The South African trajectory of this litigation ended at the SCA.¹⁰ The fact that the case never reached the Constitutional Court has important implications. It means that the Constitutional Court never had the opportunity to make authoritative pronouncements on South Africa’s obligations under the Rome Statute and meant that the Court could not resolve the question of whether sitting heads of state charged with _ius cogens_ crimes – crimes under peremptory norms of public international law – are protected by immunity before national courts.

---

² _Warrant of Arrest for Omar Hassan Ahmad Al-Bashir, Al Bashir_ (ICC-02/05-01/09) (On 4 March 2009, the Pre-Trial Chamber I issued a warrant of arrest against President Al Bashir for crimes against humanity).
³ _Second Warrant of Arrest for Omar Hassan Ahmad Al-Bashir, Al Bashir_ (ICC-02/05-01/09) (On 12 July 2010, the Pre-Trial Chamber I issued a second warrant of arrest against Al Bashir).
⁴ ‘Sudan’s President has made 74 trips across the world in the seven years he’s been wanted for war crimes’ (4 March 2016) _Quartz Africa_, available at https://qz.com/630571/sudans-president-has-made-74-trips-across-the-world-in-the-seven-years-hes-been-wanted-for-war-crimes/ (Subsequent to 2009, Al Bashir visited Malawi, Egypt, Ethiopia, Kenya, Libya, South Sudan and South Africa).
⁵ Al Bashir visited China, India, Iran, Kuwait, Saudi Arabia and Qatar (ibid).
⁶ See the following non-cooperation decisions: Decision pursuant to article 87(7) of the Rome Statute on the refusal of the Republic of Chad to comply with the cooperation requests issued by the Court with respect to the arrest and surrender of Omar Hassan Ahmad Al Bashir, 13 December 2011 (ICC-02/05-01/09-140); Decision on the Non-compliance of the Republic of Chad with the Cooperation Requests Issued by the Court Regarding the Arrest and Surrender of Omar Hassan Ahmad Al Bashir, 26 March 2013 (ICC-02/05-01/09-151); Decision on the Cooperation of the Federal Republic of Nigeria Regarding Omar Al-Bashir’s Arrest and Surrender to the Court, 5 September 2013 (ICC-02/05-01/09-159); Decision on the non-compliance by the Republic of Djibouti with the request to arrest and surrender Omar Al-Bashir to the Court and referring the matter to the United Nations Security Council and the Assembly of the State Parties to the Rome Statute, 11 July 2016 (ICC-02/05-01/09).
⁷ _Southern African Litigation Centre v Minister of Justice and Constitutional Development and Others_ [2015] ZAGPPHC 402, 2015 (5) SA 1 (GP) (‘High Court judgment’).
⁸ _Minister of Justice and Constitutional Development v Southern African Litigation Centre_ [2016] ZASCA 17, 2016 (3) SA 317 (SCA) (‘SCA judgment’).
⁹ Decision pursuant to article 87(7) of the Rome Statute on the non-compliance by South Africa with the request by the Court for the arrest and surrender of Omar Al-Bashir, 6 July 2017 (ICC-02/05-01/09-302).
¹⁰ _SCA judgment_ (note 8 above).
This article turns, first, to the flaws in the SCA’s reasoning in the *Al Bashir* case, focusing on the SCA’s thin and inadequate treatment of the norm of *ius cogens*. As will be discussed thereafter, the recourse to and application of *ius cogens* norms has been underexplored in the South African context, particularly by the Constitutional Court.

I argue that Constitutional Court jurisprudence to date has neglected the applicability of *ius cogens* norms, despite the imperative of section 232 of the Constitution. To illustrate this neglect, the cases of *Azapo*, *Wouter Basson* and the so-called *Zimbabwe Torture Docket* case will be analysed. These three Constitutional Court cases all involve the adjudication of international crimes. The *Azapo* and *Wouter Basson* cases can further be described as ‘transitional justice cases’ since they dealt with the prosecution of state-sponsored crime during apartheid.

One might blithely distinguish the *Al Bashir* case on the basis that it did not primarily involve the adjudication of the substance of international crimes, but rather the obligation on the state to cooperate with the ICC. I suggest here that *Al Bashir* is properly situated along this line of cases, precisely because the cooperation obligation is particularly compelling in the context of international crimes.

I argue that the *Al Bashir* case should have proceeded to the Constitutional Court. I do not suggest that the state’s decision to note an appeal to the Constitutional Court was correct. Rather, I am arguing that it would have been appropriate for our apex court to determine the main legal questions raised by the *Al Bashir* case, particularly the question of whether there is an *ius cogens* exception to the customary international law rule of immunity and the hierarchy between international and domestic and regional instruments.

A judgment from the Constitutional Court would have shed light on the question of the relationship between customary international law and the duty to cooperate in the prosecution of international crimes. If the case had proceeded to the Constitutional Court and the Court highlighted the *ius cogens* nature of the crimes with which Al Bashir was charged, this would have set a correct and important precedent for future cases of this kind. Such a judgment, I suggest, should have reprimanded the state for its bad faith conduct in enabling Al Bashir’s hasty departure from South Africa on 15 June 2015. Although the SCA did reprimand the state for its conduct in the 2015 Al Bashir matter, describing the conduct of government as ‘disgraceful’, a pronouncement by the Constitutional Court would have been particularly powerful and symbolic. It would also have set a more compelling precedent for future cases involving the question of whether heads of state enjoy immunity for international crimes within South Africa.

---

11 It is interesting to note that the even the Truth and Reconciliation Commission (TRC), a non-judicial body, referred to *ius cogens* norms in some of the hearings before the TRC (See legal hearing of 27–29 October 1997 on the responsibility of the judiciary, available at http://www.justice.gov.za/Trc/special/legal/legal.htm).
12 *Azanian Peoples Organisation (AZAPO) & Others v President of the Republic of South Africa & Others* [1996] ZACC 16, 1996 (4) SA 672 (CC) (*Azapo*).
13 *S v Basson* [2005] ZACC 10, 2007 (3) SA 582 (CC) (*Wouter Basson*).
15 *SCA judgment* (note 8 above) at para 7.
II THE AL BASHIR CASES

Much has been written on the Al Bashir case. The factual and litigation background to the SCA case will not be discussed at length. The most important questions dealt with by the High Court and SCA will be summarised.

A High Court case

Upon receiving confirmation of Al Bashir’s arrival on 13 June 2015, the Southern African Litigation Centre (SALC) approached the Gauteng Division of the High Court seeking the implementation of the ICC arrest warrant by means of an urgent application. When the matter was heard on 14 June the government opposed the urgent application and requested a postponement. SALC requested, and was granted, an interim order mandating the government to ensure that Al Bashir does not leave the country. On 15 June the court ruled that Al Bashir should be arrested and detained in South African pending his transfer to The Hague. At this point the counsel for the state told the court that he believed that Al Bashir had already left the country. The court then requested the state to produce an explanatory affidavit detailing how Al Bashir was allowed to leave the country despite a court order explicitly calling on government to prevent his departure.

It was clearly unlawful for government to have ignored a High Court interdict that ruled Al Bashir should not be let out of the country pending the conclusion of the application to compel government to arrest him. Section 165(5) of the Constitution states that an order or decision issued by a court binds all persons to whom and organs of state to which it applies. The court found that the Implementation Act removes head of state immunity for the crimes contained in the Rome Statute. The Rome Statute provisions on immunity, the court held, ‘means that the immunity that might otherwise have attached to President Bashir as Head of State is excluded or waived in respect of crimes and obligations under the Rome Statute’. The court further referred to a decision of the Pre-Trial Chamber of the ICC in which the Court stated that the immunities of Al Bashir ‘have been implicitly waived by the Security Council’. The State proceeded to seek leave to appeal to the SCA.

18 High Court judgment (note 7 above) at para 2.
19 Economic Freedom Fighters v Speaker of the National Assembly and Others; Democratic Alliance v Speaker of the National Assembly and Others [2016] ZACC 11, 2016 (3) SA 580 (CC) at para 1 and Department of Transport and Others v Tasima (Pty) Limited [2016] ZACC 39, 2017 (2) SA 622 (CC) at paras 179–183.
20 High Court judgment (note 7 above) at para 28.8.
22 High Court judgment (note 7 above) at paras 28.9 and 30.
B SCA case

The jurisdictional issue in the *Al Bashir* case related to the reach of South Africa’s enforcement jurisdiction.\(^{23}\) Enforcement jurisdiction refers to a state’s authority under international law to actually apply its criminal law through police and other executive action and through the courts.\(^{24}\) As the accused was present in South Africa after the commission of the alleged offences, the jurisdictional pre-conditions/nexus requirements in s 4(3)(c) of the Implementation Act were satisfied.

The SCA dealt with the relationship between articles 27 and 98 of the Rome Statute in relation to the immunity of Al Bashir. The court acknowledged that the tension between the two articles has not been authoritatively resolved.\(^{25}\) In essence, article 27 sets out the irrelevance of official capacity before the ICC.\(^{26}\) Article 98 states that the ICC shall not proceed with a request for surrender which would require the requested state to act inconsistently with its obligations under international law with regard to the state or diplomatic immunity of a person.

The SCA stated that the former Director-General at the Department of Justice and Constitutional Development, Nonkululeko Sindane, said that after South Africa agreed to host the AU Summit in June 2015 it entered into an agreement (‘the hosting agreement’) with the AU Commission relating to the organisation of the various meetings that were to take place at the summit including the 25th Assembly of the AU.\(^{27}\) Based on this agreement, Al Bashir was invited to attend the summit by the AU, and not by the government. Sindane then referred to article VIII of the hosting agreement, headed ‘Privileges and Immunities’ which states: ‘The Government shall afford the members of the Commission and Staff Members, delegates and other representatives of Inter-Governmental Organisations attending the Meetings the privileges and immunities set forth in Sections C and D, Article V and VI of the General Convention on the Privileges and Immunities of the OAU.’\(^{28}\)

The SCA held that ‘the hosting agreement did not confer any immunity on President Al Bashir and its proclamation by the Minister of International Relations and Cooperation did not serve to confer any immunity on him’.\(^{29}\) Firstly, the proclamation under s 5(3) of DIPA – the provision the government invoked – applies to organisations and their representatives. According to s 1(iv) of DIPA, “organisation” means an intergovernmental organisation of which two or more states or governments are members and which the Minister has recognised for the purposes of this Act.\(^{30}\) The SCA held that the hosting agreement provides immunity only for representatives and officials of the AU and organisations and not for those of states.\(^{31}\)

---

\(^{23}\) R O’Keefe ‘Universal Jurisdiction: Clarifying the Basic Concept’ (2004) 2(3) *Journal of International Criminal Justice* 735 (For an explanation on the difference between prescriptive and enforcement jurisdiction).

\(^{24}\) *SCA judgment* (note 8 above).

\(^{25}\) Ibid at para 60.

\(^{26}\) Article 27(2) (Sovereign immunity shall not bar the ICC from exercising jurisdiction over persons enjoying such immunity).

\(^{27}\) *SCA judgment* (note 8 above) at paras 10–11.

\(^{28}\) Ibid at para 11.

\(^{29}\) Ibid at para 47.

\(^{30}\) This refers to organisations such as the African Union or African Commission on Human and Peoples’ Rights and does not include member states or their representatives, such as heads of states.

\(^{31}\) *SCA judgment* (note 8 above) at para 42. See Tladi’s criticism of this argument (note 21 above).
It does not, therefore, provide immunity for heads of states and state delegates. Secondly, even though additional immunity can be granted to heads of states through s 7 of DIPA – also recognised in s 4(1)(a) of the Act – it would be problematic to do so in Al Bashir’s case, as it would be in contravention of the Implementation Act, which explicitly prohibits head of state immunity for Rome Statute crimes.

The SCA concluded that customary international law recognises the immunity of heads of state and other officials entitled to personal immunity from arrest and prosecution. It stated that customary international law does not recognise any exception to personal immunity even in the event of a violation of a *ius cogens* norm. The SCA, however, decided to apply and prioritise South African domestic law (as stated in the Implementation Act). It noted that although section 4 (1) of DIPA embodies the rule of the personal immunity of heads of state, that Act is subject to the Implementation Act which was enacted to domesticate South Africa’s obligations under the Rome Statute. The Court stated that ‘[t]he Implementation Act is a specific Act dealing with South Africa’s implementation of the Rome Statute. In that special area the Implementation Act must enjoy priority’.

The court further relied rather heavily on s 232 of the Constitution in justifying why the customary international law rule on personal immunities is not binding in this case. According to s 232, ‘customary international law is law in the Republic’ if it is consistent with the Constitution and Acts of Parliament. Granting Al Bashir personal immunity before South African courts would conflict with s 4(2) of the Implementation Act – the rule that head of the states do not enjoy personal immunity for international crimes. The fact that the Constitution at s 232 provides that ‘customary international law is law in the Republic unless it is inconsistent with the Constitution or an Act of Parliament’ means that *ius cogens*, as a ‘class of customary international law’ will similarly not be considered applicable law in the Republic if it is in conflict with the Constitution or an Act of Parliament. However, since *ius cogens* norms cannot be overridden by national laws, subjecting *ius cogens* norms to the test of national law is contrary to *ius cogens*. This is a far cry from what is provided for in the Constitution. In addition, requiring a *ius cogens* norm to be in line with national legislation before it can be applied is inconsistent with the fact that the principle of constitutional supremacy means that Constitutional values trump national legislation. *Ius cogens* norms play an imperative role in human rights law and the Constitution, and, as a founding value, requires the promotion of ‘human rights and freedoms’ and ‘the rule of law’.

The Implementation Act removes the bar of immunity to the surrender of an accused to the ICC where the ICC has issued an arrest warrant and made a request for cooperation. The SCA held that the conduct of the South African government in failing to take steps to arrest

---

52 Ibid.
53 Ibid at para 84.
54 Ibid at para 102.
56 The Constitutional Court recognised in *Zimbabwe Torture Docket* (note 14 above) at para 77 (Torture is a crime in the Republic as it has the status of a ‘peremptory norm of customary international law’. However, this recognition does not apply to all *ius cogens* norms as such norms are still susceptible to exclusion if they are in conflict with the Constitution or an Act of Parliament).
58 *Zimbabwe Torture Docket* (note 14 above) at para 103.
Al Bashir for surrender to the ICC was inconsistent with South Africa’s obligations under the Rome Statute and therefore unlawful.39

The SCA declined to express a view on the argument that the UN Security Council Resolution 1593 (2005) (which referred the case) implicitly waived Al Bashir’s immunity.40

III THE TREATMENT OF IUS COGENS IN THE SCA CASE

The judgment briefly refers to ius cogens and initially translates ius cogens as ‘immutable norms’. The SCA first mentions ‘peremptory norms’ in paragraph 37 when it states:

Along with torture, the international crimes of piracy, slave-trading, war crimes, crimes against humanity, genocide and apartheid require states, even in the absence of binding international-treaty law, to suppress such conduct because ‘all states have an interest as they violate values that constitute the foundation of the world public order’. Torture, whether on the scale of crimes against humanity or not, is a crime in South Africa in terms of s 232 of the Constitution because the customary international law prohibition against torture has the status of a peremptory norm.

The Court proceeds to focus on the nature of the prohibition against torture as an example of a peremptory norm which has also been domesticated in South African legislation.41 Later in the judgment, the Wallis JA refers to Weatherall:42

But the content of customary international law is not for me to determine and, like Dr Weatherall, I must conclude with regret that it would go too far to say that there is no longer any sovereign immunity for ius cogens (immutable norm) violations. Consideration of the cases and the literature goes no further than showing that Professor Dugard is correct when he says that ‘customary international law is in a state of flux in respect of immunity, both criminal and civil, for acts of violation of norms of ius cogens.43

The SCA in the Al Bashir case44 fails to situate ius cogens in the correct legal framework. By failing to situate ius cogens within the framework of treaty law – from which it originates – and by failing to mention the Vienna Convention on the Law of Treaties,45 the court does not position the norm correctly. The court’s treatment of ius cogens can therefore be described as insufficiently rigorous.

It is unfortunate that the court did not engage with the question of whether customary international law protects the immunity of heads of state and other high-ranking officials in more depth. This leaves open the possibility that government could again, in the future, rely on customary international law as a defence in favour of upholding immunity for prosecuting those who commit international crimes. Indeed, some of the more doctrinal language in the judgment could even assist those who argue that customary international law upholds immunity.46 The court essentially argued that South African domestic law trumps customary international law, which upholds immunity (a careless and unnecessary acknowledgment that

39 Ibid at para 113.
40 Ibid at para 106.
41 Ibid at paras 39–40.
43 SCA judgment (note 8 above) at para 84.
44 SCA judgment (ibid).
46 SCA judgment (note 8 above) at paras 66–68.
custom protects immunity). Instead, the SCA could have discussed the unsettled nature of the law concerning immunity for international crimes. This would have been more in line with the current position and developments in international human rights law and international criminal law.\textsuperscript{47} Even if the court did not want to make a finding on the exact position with regard to the question whether an exception exists for \textit{ius cogens} crimes, this would have made for better precedent for future cases in which the question of immunity for international crimes might arise.

The SCA did not sufficiently elaborate upon the nature of international crimes and what distinguishes international crimes from domestic crimes.\textsuperscript{48} The Heads of Arguments of the \textit{amicus curiae}, the Helen Suzman Foundation, concerned the relationship between the Constitution and the duty to investigate and prosecute international crimes:\textsuperscript{49}

The Constitution imposes special obligations on the State in respect of international crimes. This is particularly the case having regard to the nature of crimes against humanity, genocide and war crimes. Crimes of this type violate the Constitution and the State has the power and duty to detain, arrest and, in appropriate circumstances, prosecute perpetrators of these crimes. … indeed, the recognition of and protection against international crimes lies at the very core of our constitutional project.

The Heads of Arguments further stated that allowing alleged perpetrators of such crimes to avoid capture and prosecution constituted an affront to our constitutional framework.\textsuperscript{50}

\section*{IV WITHDRAWING THE APPEAL}

The state proceeded to appeal the decision of the SCA. The case was set down to be heard in the Constitutional Court on 22 November 2016. A month before the hearing, on 21 October 2016, the Minister of Justice and Constitutional Development, Michael Masutha, announced that, in essence, the SCA ‘identified the problem which needed to be addressed’ (in other words dealt with or resolved the matter) and that government will be withdrawing its appeal to the Constitutional Court.\textsuperscript{51}

It was clear to observers at the time that the Minister’s decision was strongly motivated by the state’s firm intention to withdraw from the ICC. Masutha further stated that the effect of the withdrawal from the Rome Statute as well as the repeal of the Implementation Act would complete the removal of all legal impediments inhibiting South Africa’s ability to honour its obligations relating to the granting of diplomatic immunity under customary international law as provided for under South African domestic legislation – DIPA.\textsuperscript{52} The state’s intention was illustrated by the tabling of the International Crimes Bill a little over a

\textsuperscript{47} J Dugard, G Mettraux & M du Plessis ‘Heads of State Immunities, International Crimes and President Bashir’s Visit to South Africa’ (2018) 18(4) International Criminal Law Review 577, 593 (Acknowledge the unsettled state of customary international law: ‘It cannot therefore be said…that there is a general, all-encompassing, international crimes exception to State immunities under customary international law’, speaking later about the ‘unsettled relationship’ between immunities and international crimes (ibid at 581)).

\textsuperscript{48} Akande (note 16 above).


\textsuperscript{50} Ibid at para 28.2.


\textsuperscript{52} Ibid.
year later.\textsuperscript{53} Subsequently, during June 2018, the newly-appointed Minister of International Relations, Lindiwe Sisulu, stated that government will reconsider its decision to withdraw from the ICC.\textsuperscript{54}

The legal consequences of the \textit{ius cogens} nature of the crimes with which Al Bashir was charged was thus squarely before the SCA, framed as a constitutional question. I have suggested above that the court neglected to adequately engage this question. The meagre treatment of \textit{ius cogens} norms at the SCA is not a regrettable exception, but indicative of a trend in the jurisprudence of the CC itself. I turn now to demonstrate how the Constitutional Court, too, neglected a number of necessary implications of the nature of the crimes charged in its leading cases addressing crimes proscribed by \textit{ius cogens} norms.

\section*{VI THE NATURE AND STATUS OF \textit{IUS COGENS} NORMS}

\textit{Ius cogens} was codified and recognised in article 53 of the Vienna Convention on the Law of Treaties (VCLT) where it is defined as ‘a norm accepted and recognised by the international community of states as a whole as a norm from which no derogation is permitted and which can be modified only by a subsequent norm of general international law having the same character’.\textsuperscript{55} The VCLT specifies the effect of a \textit{ius cogens} norm on a treaty: \textit{ius cogens} norms void treaties in case of conflict.\textsuperscript{56}

This includes retroactive invalidation. \textit{Ius cogens} implies a hierarchy of crimes and international crimes lie at the top of this pyramid of crimes.\textsuperscript{57} Consequently, all treaty-based rules conflicting with peremptory norms are deemed to be void and only other peremptory norms can modify the specific peremptory norm.\textsuperscript{58}

The International Law Commission (ILC) confirmed the existence, and, to an extent, recognised the importance, of \textit{ius cogens} in its preparation of the Draft Articles on the Responsibility of States for Internationally Wrongful Acts.\textsuperscript{59} The ILC, drawing on the definition provided for in, and the support of States of, the VCLT recognises that States have an interest in respecting \textit{ius cogens}.\textsuperscript{60} As to the nature of \textit{ius cogens}, the ILC provides that ‘the States have a special role’ to play in the creation of such norms and that the obligations which

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{53} Introduced in the National Assembly on 12 December 2017 (B37-2017). South Africa will however continue to be under a duty to cooperate with the ICC even if it withdraws from the ICC (N Pillay and A Mudukuti ‘South Africa and the ICC: Dismantling the International Criminal Justice System to Protect One Individual’ (19 June 2018) \textit{Daily Maverick}, available at https://www.dailymaverick.co.za/article/2018-06-19-south-africa-and-the-icc-dismantling-the-international-criminal-justice-system-to-protect-one-individual/#W0saqtUzBq8)).
\item \textsuperscript{55} VCLT, art 53.
\item \textsuperscript{56} VCLT, art 64 (If a new peremptory norm of general international law emerges, any existing treaty which is in conflict with that norm becomes void and terminates).
\item \textsuperscript{58} VCLT, art 53.
\item \textsuperscript{59} These articles were adopted by the ILC in 2001 (see note 61 below).
\item \textsuperscript{60} G Arangio-Ruiz, Special Rapporteur on State Responsibility ‘Fourth Report on State Responsibility’ (1992) II(1) \textit{Yearbook of International Law Commission} 33.
\end{itemize}
\end{footnotesize}
**ius cogens** impose upon States ‘affect the vital interest of the international community as a whole’. Such obligations may even attract a higher responsibility for States than those ‘applied to other internationally wrongful acts’. The notion that **ius cogens** norms are not static and the reason why there is no closed list is bolstered by the fact that article 64 of the VCLT recognises the possibility of the creation of new **ius cogens** norms. However, the ILC goes further than the VCLT by recognising the assertion that non-derogation from **ius cogens** norms includes all legal acts as opposed to only treaty obligations.

**ius cogens** norms are closely related to the concept of non-derogability. The Inter-American Court of Human Rights has stated that non-derogable treaty rights constitute an important starting point when identifying **ius cogens** norms. The VCLT does not include examples of norms of **ius cogens**. Instead it allows States, international judicial bodies and scholars to establish which norms meet the requirements of article 53 of the VCLT, which leaves it to the ‘international community as a whole’ to identify those international law norms belonging to **ius cogens**. According to de Wet this means that a particular norm is first recognised as customary international law, where after the international community of states as a whole further agrees that it is a norm from which no derogation is permitted. A peremptory norm would therefore be subject to ‘double acceptance’ by the international community of states as a whole.

The ILC has described the following norms as those that are most frequently cited as having **ius cogens** status: (a) the prohibition of aggressive use of force; (b) the right to self-defence; (c) the prohibition of genocide; (d) the prohibition of torture; (e) crimes against humanity; (f) the prohibition of slavery and slave trade; (g) the prohibition of piracy; (h) the prohibition of racial discrimination and apartheid; and (i) the prohibition of hostilities directed at civilian population.

At its root **ius cogens** also draws upon elements of natural law and the two concepts share some overt similarities, notably as being ostensibly ‘moral’ doctrines. Orakhelashvili is of the

---

66 Ibid.
68 The definition of **ius cogens** in the VCLT was influenced, in particular, by the work of Albert Verdross who was strongly influenced by natural law. A Verdross ‘Jus Dispositivum and jus Cogens in International Law’ (1966) 60 American Journal of International Law 56.
view that ‘[m]orality can arguably itself explain a norm’s peremptory character’.69 Like natural law, *ius cogens* is binding regardless of consent. The similarities between natural law and *ius cogens* were observed during the drafting of the VCLT (at times the delegates used the terms interchangeably).70 Although most scholars tie the origin of *ius cogens* to natural law, some scholars of a more positivist inclination argue that *ius cogens* emanated from the will of states as expressed in treaties or in custom.71

*Ius cogens* remains difficult to define. Georges Abi-Saab captured the elusive nature of the concept when he said that even if the normative category of *ius cogens* were to be an ‘empty box, the category was still useful; for without the box, it cannot be filled’.72 Tied to the difficulty of definition is the thorny issue of determining the content of *ius cogens*. There are two dominant views: Some scholars hold the view that there is a large measure of international consensus on the basic norms that qualify as *ius cogens* norms: the prohibition against slavery, torture, racial discrimination, self-determination, piracy and the prohibition of the use of force.73 The crime of apartheid is also generally included in this category of norms. A second school of thought believes that the content of *ius cogens* is vague and indeterminate.74 These scholars argue that the uncertain content of *ius cogens* pose a threat to the very viability of *ius cogens*.75

It has to be acknowledged that *ius cogens* remains both certain and uncertain. Whereas to claim that the norm is entirely certain would contradict the very nature of *ius cogens*, however the uncertainty has been overemphasised. One scholar went as far as claiming that, in determining when a norm reaches the status of *ius cogens*, ‘it appears that judges and scholars simply consult their own consciences’.76 The VCLT states that in terms of the Convention, *ius cogens* is a ‘norm of general international law … accepted and recognized by the international community of States as a whole’.77 This language makes an explicit link between the nature of the norm (whether peremptory or not) and the question of standing in public international law (whether only one affected state or potentially any state may invoke a breach of a norm; in other words, whether the obligation is owed merely *inter partes* or *erga omnes*). While the point at which a norm attains the status of *ius cogens* does remain unclear, the determination of *ius cogens* is not as subjective as being merely a matter of consulting one’s conscience. There is a substantial measure of consensus on a number of *ius cogens* norms.

71 Ibid at para 53.
77 VCLT, art 53.
Whereas the content of *erga omnes* is fairly well established in the case law of the ICJ, the notion of *ius cogens* has not received the same attention by the ICJ and has been cited relatively rarely. It only entered ICJ case law to the extent it deserves relatively recently.

VII EXCEPTION TO IMMUNITY

There is as yet little evidence of state practice or *opinio juris* that sitting heads of State can be tried for international crimes. As Kiyani points out, the cases that establish this rule usually suffer from one or both of the following flaws: the individuals concerned, such as Muammar Gadaffi and Charles Taylor, were no longer sitting heads of State at the relevant time, and therefore lacked personal immunity; or, they had been transferred, with the consent or co-operation of the State, to the court in question, which amounts to a waiver of the immunity by the State concerned. According to Kiyani these reasons fit with ordinary understandings of head of State immunity, and show no evidence of an exception to those rules. Claus Kreß is similarly of the view that a rule of custom is ‘admittedly not (yet) firmly entrenched and fortified’ in this context.

Whereas there is more *opinio juris* to rely on in support of the exception than state practice, *opinio juris* is still not sufficiently strong and cannot bear the entire burden of proving that the exception had crystallised into custom. Dapo Akande, for example, strongly doubts the existence of sufficient *opinio juris*.

VIII *IUS COGENS* IN INTERNATIONAL JURISPRUDENCE

The ICJ, despite repeated reference to ‘general and fundamental principles which lie beyond contractual treaty-relations’ in its reasoning, has been reluctant to include the concept of *ius cogens* in its judgments. The court was initially equally reluctant to refer to *erga omnes* norms but eventually started to refer to *erga omnes* in the *Barcelona Traction* case. In this case the ICJ distinguished between obligations a state has towards other states and obligations a state has towards the international community.

As Hernandez points out, peremptory norms are not new to the ICJ. As early as 1934, the Permanent Court of International Justice (PCIJ), in the separate opinion of Schücking J, made reference to the concept of *ius cogens* in the *Oscar Chinn* case and linked it expressly to international public morality. Schücking J stated that the PCIJ would never ‘apply a Convention the terms of which were contrary to public morality’. He implied that public

---


80 *Prosecutor v Omar Hassan Al Bashir* ICC/02-05-01/09 OA2 (18 June 2018) para 8 (Written observations of amicus curiae, Prof Claus Kress).

81 Akande (note 16 above).

82 *Barcelona Traction* (note 78 above).

83 Ibid at para 33.


85 Ibid. *Oscar Chinn case (Britain v Belgium)* (1934) PCIJ Series A/B No 63 (*’Oscar Chinn’*) (Separate opinion, Schücking J) 149, 150.

86 Ibid.
morality might constitute *ius cogens*. By making this link to public morality, Schücking J acknowledges the natural law roots of *ius cogens* discussed above. His statement suggested that every legal system can have a set of norms which are beyond the contractual disposition of the contracting parties.

Why has the ICJ been this reluctant to rely on *ius cogens*? One explanation entails a sources-of-law problem. Applying *ius cogens* norms presents challenges to the international judicial function. It requires judges to address the effects of *ius cogens* norms without being able to test the validity of the norms. Hernandez explains why this is the case: ‘[t]o do so sets international law decisively on a path away from its classical foundations as a system regulating relations between completely sovereign States and erected purely through their willed consent’.  

The court has on occasion referred to *ius cogens* without using the term *ius cogens*. In *Diallo* for example, the ICJ refrained from using the terms ‘peremptory’ or *ius cogens*, but concluded that the ‘prohibition of inhuman and degrading treatment is among the rules of general international law which are binding on States in all circumstances, even apart from any treaty commitments’. In the *Nicaragua*, *Oil Platforms* and *Armed Activities in the Congo* judgments, the imperative character of the prohibition of the use of force was given no legal effect.

In the *Congo v Rwanda* case, Dugard J applauds the decision of the ICJ to explicitly recognise and accept the notion of *ius cogens*. Dugard J writes that such acceptance is long overdue. Whereas *ius cogens* has long been accepted by international law academics, international courts have been slow to award *ius cogens* its rightful status. In his separate opinion Dugard J discusses the role of *ius cogens* in the judicial decision-making process and concludes that, in the exercise of its judicial function, when choosing between competing sources the court should choose *ius cogens*. To Dugard J, the *South West Africa* cases presented the perfect opportunity to apply *ius cogens*. He argues that the *East Timor* decision and the *Arrest*

---

88 Hernandez (note 84 above) at 40.
90 *Case Concerning Military and Paramilitary Activities and Against Nicaragua (Nicaragua v United States of America)* Merit Judgment, ICJ Reports (1986) 14 (‘Nicaragua’) para 190.
91 *Case Concerning Oil Platforms (Islamic Republic of Iran v United States of America)* Judgment ICJ Reports (2003) 161 (‘Oil Platforms’).
93 *Nicaragua* (note 90 above) at para 190 (The Court may have mentioned *ius cogens* twice, but only through quoting directly from the United States’ Counter-Memorial). The same approach was taken in *Oil Platforms* (note 91 above) at para 81, this time using the United States’ Rejoinder.
95 Ibid at 89, para 8.
97 *Congo v Rwanda* (note 94 above) at 89, para 10.
98 *South West Africa (Ethiopia v South Africa; Liberia v South Africa)* Judgment ICJ Reports (1966) 6.
99 *Congo v Rwanda* (note 94 above) at 88, para 4.
100 *Case Concerning East Timor (Portugal v Australia)* Judgment ICJ Reports (1995) 90 (‘East Timor’).
Warrant case\textsuperscript{101} would also have been suitable cases in which to invoke \textit{ius cogens}.\textsuperscript{102} In these cases the ICJ was faced with competing principles, precedent and state practice and preferred not to choose that solution which gave effect to a norm of \textit{ius cogens}.\textsuperscript{103} In contrast to 	extit{Congo v Rwanda}, the \textit{Arrest Warrant} case can be described as a low point in the ICJ’s treatment of \textit{ius cogens}. The ICJ held, without express reference to \textit{ius cogens}, that the fact that a Minister of Foreign Affairs was accused of a violation of rules which undoubtedly possess the character of \textit{ius cogens} did not remove the immunity he enjoys under customary international law.\textsuperscript{104} It is largely because of the retrogressive decision in \textit{Arrest Warrant} that the law remains unsettled.

Because of his position in \textit{Congo v Rwanda}, Dugard has been credited, both as a judge and academic, with championing a doctrinal move towards the recognition of \textit{ius cogens}.\textsuperscript{105} But outside the confines of ICJ jurisprudence, the position regarding the relationship between immunity and \textit{ius cogens} is still not entirely acknowledged by international and domestic courts. In the ground-breaking \textit{Pinochet} judgment for example, only a minority of judges, Judge Philips and Judge Millet, opined that systematic torture was an international crime for which there could be no immunity.\textsuperscript{106} In the unfortunate \textit{Al-Adsani} case before the European Court of Human Rights, it was held that international law as it stood provided no basis for concluding that a state no longer enjoys immunity for civil suits in the courts of another state where acts of torture had been alleged.\textsuperscript{107}

Another example from the International Court of Justice is instructive. The dissenting opinions in the \textit{Jurisdictional Immunities} case by Judges Trindade, Yusuf and Judge \textit{ad hoc} Gaja all present strong arguments in favour of the \textit{ius cogens} trumping the law of immunity. Judge Trindade wrote a powerful dissent in which he argued that the majority decision did not serve justice.\textsuperscript{108} He essentially upheld the primacy and hierarchical superiority of \textit{ius cogens}. Trindade J describes the distinction between procedure and substance in the context of the immunity versus \textit{ius cogens} debate as ‘formalist’.\textsuperscript{109} He writes that this distinction deprives \textit{ius cogens} of its effects and legal consequences.\textsuperscript{110} In discussing the question whether an exception exists to the general applicability of immunity in cases of breaches of \textit{ius cogens} norms, Judge Gaja refers to a minority opinion in the European Court of Human Rights in \textit{Al-Adsani}\textsuperscript{111} and to a number of judgments by the Italian \textit{Corte di Cassazione}, especially those in \textit{Ferrini}.\textsuperscript{112}

\begin{flushright}
\textsuperscript{101} \textit{Case Concerning the Arrest Warrant of 11 April 2000 (Democratic Republic of the Congo v Belgium)} International Court of Justice (14 February 2002), available at https://www.refworld.org/cases,ICJ,3c6cd39b4.html.
\textsuperscript{102} \textit{Congo v Rwanda} (note 94 above) at 89, para 11.
\textsuperscript{103} Ibid.
\textsuperscript{104} \textit{Arrest Warrant} (note 101 above) at paras 54–58.
\textsuperscript{106} \textit{R v Bow Street Metropolitan Stipendiary Magistrate: Ex Parte Pinochet Ugarte} (No. 3) [1999] 2 All ER 97 (HL) (‘\textit{Pinochet}’).
\textsuperscript{107} \textit{Al-Adsani} (note 1 above) at paras 101–102.
\textsuperscript{108} \textit{Jurisdictional Immunities of the State (Germany v Italy: Greece Intervening)} Judgment, ICJ Reports (2012) 99 (‘\textit{Jurisdictional Immunities}’) at 179, para 1.
\textsuperscript{109} Ibid at para 315.
\textsuperscript{110} Ibid.
\textsuperscript{111} , Ibid at 320, para 11 (Dissenting opinion of Judge \textit{ad hoc} Gaja). See also \textit{Al-Adsani} (note 1 above).
\end{flushright}
and in Milde.\textsuperscript{113} He also referred to a decision by the French Cour de cassation in La Reunion Aerienn\textsuperscript{e}\textsuperscript{114} which pointed to the existence of a restriction of immunity when a claim concerns international crimes of a \textit{ius cogens} nature. The jurisprudence cited by Judge Gaja points to the unsettled nature of the law of immunities. Whereas it cannot be said definitively that a rule has crystallised that immunities do not apply to \textit{ius cogens} crimes, it can also not be said that the law currently supports the position that \textit{ius cogens} norms do not trump immunity.

It is clear that the \textit{Al Bashir} case could have benefitted from a clearer and more nuanced discussion of the current state of international law with regard to the question of immunity for international crimes that rise to the level of \textit{ius cogens}. These legal issues were squarely before the court. But the \textit{Al Bashir} case is not unique in terms of the absence of a principled discussion of \textit{ius cogens}. The South African courts have only rarely discussed \textit{ius cogens}. In cases on international crimes, most notably the \textit{Azapo} case, the Wouter Basson case, \textit{ius cogens} was entirely neglected. The Zimbabwe Torture Docket case, the most progressive case on international criminal law heard by the Constitutional Court to date, did refer to \textit{ius cogens} extensively.

IV  \textbf{IUS COGENS IN SOUTH AFRICAN LAW AND JURISPRUDENCE}

A  \textit{Azapo} case

The \textit{Azapo} case was the first Constitutional Court case that involved the question of the status of international crimes after the political transition. This case involved a challenge to the amnesty provisions in s 20(7) of the Promotion of National Unity and Reconciliation Act 34 of 1995 (Reconciliation Act). Steve Biko's widow and other family members of apartheid victims argued that the amnesty provisions in the Act violated their constitutional right to access to justice as recognised in s 34 of the Constitution. The Constitutional Court held that the epilogue to the Constitution trumped s 34 of the Constitution and that s 20 (7) of the Reconciliation Act authorising criminal and civil amnesty was therefore constitutional.\textsuperscript{115}

It has been argued that laws which give amnesty to alleged perpetrators of international crimes are in violation of \textit{ius cogens}.\textsuperscript{116} The \textit{Azapo} judgment however did not mention \textit{ius cogens} norms. The judgment does refer to the tensions between justice for those wronged during conflict, on the one hand, and the consolidation of the transition to a nascent democracy but the judgment failed to take international law, particularly international humanitarian law, seriously.\textsuperscript{117}

\begin{footnotesize}
\begin{enumerate}
\item[113] Germany v Milde (Max Joseph) (2009) 92 Riv Dir Int 618, ILDC 1224 (IT 2009) (‘Milde’).
\item[114] La Reunion Aerienn\textsuperscript{e} and Ors v Libya Arab Jamahiriya ILDC 1770 (FR 2011) (‘La Reunion Aerienn\textsuperscript{e}’).
\item[116] F Muhammadin ‘The Legality of Amnesty Laws for Ius Cogens Crimes under the Norms of ius cogens’ (10 December 1998) undergraduate thesis, submitted and defended before the Board of Examiners, Department of International Law, Faculty of Law, University Gadjah Mada (Indonesia) as required to obtain the Sarjana Hukum (LLB equivalent) degree, available at https://www.academia.edu/8573163/The_Legality_of_Amnesty_Laws_for_Ius_Cogens_Crimes_under_the_Norms_of_Ius_Cogens, 5. See also Prosecutor v Furundzija (Judgment) ICTY Trial Chamber IT-95-17/1.
\item[117] Azapo (note 12 above) at para 31.
\end{enumerate}
\end{footnotesize}
The judgment, otherwise eloquently pronounced by Mahomed J, not only failed to make any mention of *ius cogens* or of the fact that apartheid constituted a crime against humanity.\(^{118}\) By limiting its discussion to war crimes, the Court failed to appreciate the fact that apartheid is – and was, at least from 1966 – a crime against humanity.\(^{119}\) The Court further found that the 1949 Geneva Conventions and the Additional Protocols (in their entirety) did not apply to South Africa.\(^{120}\) Crucially, the Court did not appreciate that the Second Additional Protocol to the Geneva Conventions (applied to national liberation movements) was drafted specifically with the apartheid regime in mind. In any event, there is considerable academic authority for the proposition that the Geneva Conventions constituted customary international law before 1998 (the date of the signing of the Rome Statute) in large part because of the near-universal ratification of the Conventions.\(^{121}\) Whereas apartheid crimes cannot be prosecuted under the Rome Statute (because of its limited temporal jurisdiction) it is clear that such crimes can (and indeed must) be prosecuted under the Geneva Conventions, or under customary international law.

Dugard suggests that AZAPO was disappointing because it failed to address whether conventional and customary international law oblige a successor regime to punish the officials and agents of a prior regime for violations of international law.\(^{122}\) He wrote that both treaty and customary international law oblige a successor regime to punish members of the prior regime for acts that constitute crimes under international law. Dugard wrote:

> The judgment does not, however, show concern for the international legal order that condemned apartheid as a crime against humanity, that served as a standard by which the laws of apartheid were measured in the years without hope. This was a judgment that called for the broad brush of history, for an exposition of why apartheid was judged by the international community to be a crime against humanity, for an examination of the experience of other societies that have emerged from darkness, for an explanation of the reason why international law does not compel a society bent on reconciliation to prosecute those who have committed the most heinous crimes in the name of the state.\(^{123}\)

The Constitutional Court failed to recognise the fact that international crimes are of a fundamentally different nature than domestic crimes. What differentiates murder as a crime against humanity from a crime under South African domestic law is both its distinctive contextual element – ‘committed as part of a widespread or systematic attack directed against any civilian population, with knowledge of the attack’\(^{124}\) – and the corollary that permissive universal jurisdiction would attach to murder as a crime against humanity.\(^{125}\)

By ignoring the norm of *ius cogens*, the Court further failed to take account of the normative evolution of international law and failed to integrate this evolution into South African Constitutional law.


\(^{120}\) *Azapo* (note 12 above) at para 29.

\(^{121}\) du Plessis (note 119 above) at 3; and Canada’s Crimes Against Humanity and War Crimes Act S.C. 2000, c.24. See also the Final Report of the Truth and Reconciliation Commission.


\(^{123}\) Dugard (ibid) at 262.

\(^{124}\) Rome Statute, art 7.

\(^{125}\) *Zimbabwe Torture Docket* (note14 above) at para 40.
B Wouter Basson case

The Wouter Basson case was first case regarding the prosecution of crimes committed by an individual on the instructions of the apartheid government to reach the Constitutional Court. Wouter Basson, a cardiologist colloquially referred to as ‘Dr Death,’ was indicted in 1999 in the Pretoria High Court on 67 counts, with charges ranging from 229 murders and conspiracy to murder as well as the manufacturing, possessing, and dealing of drugs and fraud for personal enrichment totalling R36 million. As the head of the South African Defence Forces (SANDF) biological and chemical program called ‘Project Coast’ from 1981 to 1995, Basson oversaw the planning and execution of a number of mass murders committed by clandestine units of the South African National Defence Forces. Basson was alleged to have been involved in conduct amounting to war crimes committed outside South African territory during the Namibian border war. By drafting the conspiracy charges with reference to domestic legislation, the State prosecutors implicitly framed the egregious acts of violence as instances of ordinary criminality. When the case was heard by the High Court, the court not only failed to mention *ius cogens* but failed to discuss certain fundamental principles of international law. When the case was appealed to the SCA, the SCA refused to hear the merit of the legal question of whether the charges were correctly quashed because the application was fundamentally defective in various procedural respects.

The Constitutional Court application concerned the three central issues of (i) bias of the trial judge (ii) admissibility of the bail record and (iii) the quashing of six charges. The Court considered it crucial to consider the seriousness of the charges in light of South Africa’s international obligations regarding the maintenance of fundamental principles of international humanitarian law. In light of the gravity of the offences the question arose of whether South Africa had an obligation to prosecute under international law. The Constitutional Court similarly failed to refer to *ius cogens*. This was partly as a result of the Court’s insistence to frame the case as a domestic case rather than a case that was based on the principle of universal jurisdiction or flowing from South Africa’s obligations at customary or conventional international law. The Constitutional Court justices implicitly framed Basson’s conduct as analogous to instances of ordinary criminality such as murder, rape, and assault. Although the Court’s judgment alludes briefly to the state’s obligations under international law to prosecute international crimes such crimes against humanity and war crimes, these obligations are framed as secondary to those imposed by domestic law. While the state, on appeal, argued that Basson’s conduct as crimes against humanity and war crimes were criminalised under

---

126 Wouter Basson (note 13 above). The first case brought against apartheid era violators and that ended in acquittals in 1996 was *S v Msane and 19 others* (Unreported judgment case no CC1/96 heard in the Durban and Coast Local Division of the Supreme Court of South Africa).

127 According to Advocates Ackermann and Pretorius, Basson was charged with conspiracy because the actual conspiring took place in Pretoria, thus falling under the jurisdiction of South African courts and the country’s domestic law.


130 Wouter Basson (note 13 above) at para 171.

131 A Raleigh ‘Charging decisions, legal framing and transitional justice: the prosecution of Wouter Basson’ (2019) 35(2) SAJHR 194 (On the reasons why the case was framed in terms of domestic law).
customary international law at the time they were committed, the dominant frame of reference remains that of domestic law.

In view of the seriousness of the charges, it would have been more appropriate and in accordance with the expectations of the international community to apply universal jurisdiction to Basson’s alleged conduct committed within the territory of Namibia and outside South African territorial jurisdiction.\textsuperscript{132} The existence of jurisdictional links (both territorial and personal) does not exclude the possibility of referring to universal jurisdiction. Applying universal jurisdiction in this case could have lent the appropriate status to the alleged crimes. It can also be asked whether international criminal law will develop as desired if universal jurisdiction is considered as a subsidiary form of jurisdiction: a form of jurisdiction a court will only resort to in the absence of all other jurisdictional links. Such a practice could also lend substance to the fear expressed in \textit{Tadic} that ‘human nature being what it is, there would be a perennial danger of international crimes being characterized as “ordinary crimes”’.\textsuperscript{133}

Although \textit{ius cogens} has been under-recognised by the Constitutional Court, the gravity and status of genocide as a \textit{ius cogens} norm has been recognised in academic literature on the \textit{Basson} case.\textsuperscript{134} Although the Constitutional Court case focused on charges of conspiracy and not on the fertility ‘research’ conducted by Basson, the Court recognised the existence of a duty to prosecute.\textsuperscript{135} The Court held that the National Prosecuting Authority is under an international obligation to prosecute crimes committed during the apartheid era.\textsuperscript{136}

Whereas ordinary criminality falls under the domain of the domestic law frame, the international law frame concerns itself with what Drumbl has termed ‘extraordinary international criminality,’\textsuperscript{137} a designation that suggests that such crimes exist as substantively different forms of criminality than ordinary common crimes such as murder and rape.

\textbf{C \hfill Zimbabwe Torture Docket case}\

The so-called ‘Zimbabwe Torture Docket case’\textsuperscript{138} involved the duty under domestic and international law to investigate international crimes. The case involved allegations of widespread torture, amounting to crimes against humanity, committed by Zimbabwean officials in Zimbabwe. The case concerned the South African Police Service’s (SAPS) responsibilities under international and domestic law to investigate international crimes committed in Zimbabwe, which is not a state party to the ICC.\textsuperscript{139}

The case raised a number of interesting international and domestic legal questions regarding the exercise of universal jurisdiction. The main questions include: the legality of

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{132} However, basing the case on international law would not necessarily have resulted in a conviction (ibid).
\item \textsuperscript{133} \textit{Prosecutor v Dusko Tadic} Appeals Chamber IT-94-1-AR72 Decision on the Defence Motion for Interlocutory Appeal on Jurisdiction 2/10/1995 para 58.
\item \textsuperscript{134} M Jackson ‘A Conspiracy to Commit Genocide: Anti-Fertility Research in Apartheid’s Chemical and Biological Weapons Programme’ (2015) 13(5) \textit{Journal of International Criminal Justice} 993.
\item \textsuperscript{135} Wouter Basson (note 13 above) at paras 31‒37.
\item \textsuperscript{136} Ibid at para 37. H van der Merwe \textit{The Transformative Value of International Criminal Law} (2012) 265,(Thesis presented in fulfillment of the requirements for the degree of Legum Doctor (LLD) in the Faculty of Law at Stellenbosch University).
\item \textsuperscript{137} M Drumbl \textit{Atrocity, Punishment and International Law} (2007) 4.
\item \textsuperscript{138} \textit{Zimbabwe Torture Docket} (note 14 above).
\end{itemize}
\end{footnotesize}
universal jurisdiction in absentia under international law; the correct interpretation of the Implementation Act’s ‘presence’ requirement, the relevance of the fact that Zimbabwe is not a party to the Rome Statute as well as the question of whether there is an obligation to prosecute international crimes under international law or domestic Constitutional law.\textsuperscript{140}

The Court’s key findings were, firstly, that South Africa can exercise universal jurisdiction over international crimes such as torture under both international and domestic law; secondly, the presence of the suspect in South Africa was not required under international or domestic law in order to begin an investigation; and, thirdly, that South Africa was under an obligation to investigate such crimes under international law – and that, in terms of domestic law, such investigation is to be ‘discharged through … law-enforcement agencies’ (i.e. the SAPS).\textsuperscript{141} According to du Plessis the Court exercised a ‘robust’ form of universal jurisdiction – in terms of which the presence of the accused on a state’s territory is not required.\textsuperscript{142} The court further found that domestic law also does not require such presence in the early stages of an investigation.\textsuperscript{143}

After discussing the various ways in which a court can ground jurisdiction for criminal offences (including the possibility of exercising universal jurisdiction) the Court recited the \textit{ius cogens} nature of torture under international law and the universal condemnation it attracts.\textsuperscript{144} The Court stated that states are obliged to prosecute piracy, slave trading, war crimes and crimes against humanity, genocide and apartheid even in the absence of treaty obligations.\textsuperscript{145} The Court further stated that torture was a crime under s 232 of the Constitution.\textsuperscript{146} The Court mentioned that the obligation to prosecute torture was included in the Prevention and Combating of Torture of Persons Act.\textsuperscript{147}

\textit{ius cogens} is first referred to in footnote 2 of the judgment which reads:

\begin{quote}
A state’s duty to prevent impunity, which can be defined as the exemption from punishment, is particularly pronounced with respect to those norms, such as the prohibition on torture, that are widely considered peremptory and therefore non-derogable – even in times of war or national emergency – and which, if unpunished, engender feelings of lawlessness, disempower ordinary citizens and offend against the human conscience.\textsuperscript{148}
\end{quote}

\textit{ius cogens} features again at paragraph 35 where the Court states: ‘Coupled with treaty obligations, the ban on torture has the customary international law status of a peremptory norm from which no derogation is permitted.’

Finally, \textit{ius cogens} appears in the concluding section of the judgment where the Court provides ‘[the crime of torture] law in the Republic in terms of section 232 of the Constitution due to its status as a peremptory norm of customary international law.’\textsuperscript{149} The conclusion of the judgment is particularly sensitive to South Africa’s obligations as part of the ‘community

\begin{footnotes}
\item[141] \textit{Zimbabwe Torture Docket} (note 14 above) at para 50.
\item[142] Du Plessis (note 139 above).
\item[143] \textit{Zimbabwe Torture Docket} (note 14 above) at para 49.
\item[144] Ibid at para 36.
\item[145] Ibid at para 37.
\item[146] Ibid.
\item[147] Ibid at para 38.
\item[149] \textit{Zimbabwe Torture Docket} (note 14 above) at para 77.
\end{footnotes}
of nations’.\textsuperscript{150} It is therefore particularly unfortunate that the case failed when it came to implementation – no investigation followed. The Zimbabwe Torture docket case nevertheless clearly represents a progression in the Constitutional Court’s treatment of international crimes. It can be described as the most progressive domestic case on international criminal law not just in South Africa but also arguably in Africa.

V CONCLUSION

Due to South Africa’s isolation during apartheid, the South African judiciary has only relatively recently started to seriously engage with international law. The neglect of international law during the Apartheid era means that South African lawyers were traditionally not well schooled in the application of international law (especially customary international law) and for a long time international law has been considered too unfamiliar and exotic to warrant reliance.\textsuperscript{151} Whereas the Constitutional Court initially showed an awkwardness in dealing with this international law doctrine (as evident from the Azapo and Wouter Basson judgments), the Court showed much greater sensitivity to the principles underlying international crimes in the Zimbabwe Torture Docket case. This indicates an important evolution of South African constitutional law which will hopefully continue.

Gevers describes South Africa’s relationship with international criminal law as ‘complex’ and ‘schizophrenic’.\textsuperscript{152} It is clear that the spectre of state sovereignty still hampers the development of the new international law. It can be argued that the ‘spectre of sovereignty’ manifests itself in the SCA and the Constitutional Court’s reluctance to rely on the \emph{ius cogens} nature of certain international crimes. The Constitutional Court, in the Zimbabwe Torture Docket case, has however taken a step forward by engaging more closely with \emph{ius cogens} than the SCA. In the Al Bashir case, as I demonstrated above, the SCA treated \emph{ius cogens} as an afterthought and not as a central part of its decision.

In applying \emph{ius cogens}, South African courts would do well to consider comparative jurisprudence. Switzerland, for example, has a commendable way of protecting \emph{ius cogens} norms of international law within the domestic legal order. By providing constitutional recognition of \emph{ius cogens} norms, the revised Swiss Federal Constitution of 1999 ensures the superior status of these norms.\textsuperscript{153} The 1999 Constitution explicitly states that no People’s Initiative (referendum) aimed at constitutional amendment may be in conflict with the norms of \emph{ius cogens}.\textsuperscript{154} If \emph{ius cogens} norms were given constitutional status in South Africa by including these norms in the South African Constitution judges would have to consider \emph{ius cogens} whenever they are confronted with a case involving such norms. It would no longer be possible for judges to ignore or downplay \emph{ius cogens} norms as occurred when the Al Bashir case reached the SCA.

The Constitutional Court has yet to address and acknowledge the notion of hierarchically superior norms in international law and affirm the superiority of such norms to domestic

\textsuperscript{150} Ibid at para 80.
\textsuperscript{151} Swart (note 128 above) at 213.
\textsuperscript{154} Ibid.
norms. It is ironic that the Constitutional Court, itself tasked with upholding a set of superior norms, has not acknowledged the *ius cogens* nature of the core international crimes. The Constitutional Court, which lies at the apex of the South African legal system, failed to appropriately recognise a norm that lies at the apex of the international legal order and that is hierarchically superior to all international legal norms.
**Constitutional Court Review**

**Instructions for Authors**

The *Constitutional Court Review* (CCR) is an international journal of record that tracks the work of the Constitutional Court of South Africa. The long essays, replies, articles and case comments use recent decisions to navigate more general currents in the Court’s jurisprudence. The Journal follows a strict double-blind, peer-reviewed editorial process. The CCR invites contributions from outstanding scholars but also considers unsolicited submissions that fit with the aims and scope of the Journal. It is published annually.

Articles are published Open Access under a Creative Commons Attribution Licence (CC BY).

### I. EDITORIAL POLICY

Submission of a manuscript implies that the material has not previously been published, nor is being considered for publication elsewhere. Contributions are accepted on the understanding that the authors have the authority for publication. Contributions must conform to the principles outlined in [Ethical considerations in research publication](https://www.nisc.co.za/products/97/journals/constitutional-court-review). The Journal has a policy of anonymous (double-blind) peer review. Authors’ names are withheld from referees; authors and the editors must ensure that any identifying material is removed from the manuscript. The Editor reserves the right to revise the final draft of the manuscript to conform to editorial (house style) requirements.

Articles accepted for publication will attract an article processing charge (APC) of ZAR 12 000 (excl. VAT). The Journal has an APC waiver policy for authors whose funding arrangements are inadequate to cover the amount of the APC. The APC waiver-policy and procedures for applying for a waiver can be found at [https://www.nisc.co.za/openaccess](https://www.nisc.co.za/openaccess). Applications, which are considered on a case-by-case basis, must be made prior to submission via email to journals@nisc.co.za. Note that the Journal has an annual quota for APC waivers. It may not be possible to accommodate all APC waiver requests made during the year.

### I. CONTRIBUTION FORMAT & LENGTH

As a general rule, the *Constitutional Court Review* accepts submissions in the following formats:

- **Lead Essays and Responses:** Essay length will vary depending upon the nature of the contribution. We have had lead essays that range from 65 to 80 pages. Thus, these lead essays – commissioned as such – may run over 30 000 words. Responses to lead essays – commissioned as such – may range from 10 000 to 25 000 words. Replies have been, on occasion, longer than the Lead Essay.

- **Articles:** Articles, commissioned as such or received by open call, should fall within a range of 12 000 to 25 000 words. As with all pieces, exceptions can and will be made.

- **Comments:** The distinction between articles and case comments can be somewhat artificial. Case comments will often have a range and depth that outstrip articles (at which point we often call them articles.) However, for pieces that narrow their focus to a single case and its more limited ramifications, we would expect between 6 000 to 10 000 words.
Where exceptions are required, they will be negotiated between editors and authors. Consult recent copies of the journal at <https://www.nisc.co.za/products/97/journals/constitutional-court-review> for examples of each type of contribution.

III. MANUSCRIPT PRESENTATION & SUBMISSION

Manuscripts should be emailed to the Editor-in-Chief, Professor Stu Woolman at: <stuart.woolman@wits.ac.za>. Manuscripts should be in English and prepared in MS-Word format. Submissions should include a title page (as a separate file) with the following information:

- **Title.** The title should be short and descriptive, bearing in mind the need for discoverability using standard search terms.
- **Abstract** of 200 to 300 words
- **Keywords.** These four to six terms can be single words or phrases of more than one word. However, phrases tend to be less helpful as search terms in isolating groups of similar articles because the phrases are often unique (to that particular piece). The keywords facilitate discovery of the article and so should be chosen with care. Keywords should not repeat words/terms that are already in the title, since the title will already be parsed by search engines
- **Author details.** Full name, title, institutional (or other) affiliation(s), and email address
- **Acknowledgements.** (e.g. ‘I thank X, Y, Z, the editors and anonymous reviewers for their helpful comments on this piece.’)

The body of the article should be in a separate document. Please remove names and other identifiers to facilitate anonymous peer review. Prepare the manuscript according to the format and style conventions below. Manuscripts that do not conform to the Journal’s style and format conventions may be returned to the author for remedy without further evaluation. In general, authors should strive to present their arguments clearly and should avoid repetition and padding. We ask authors to be ruthless with their own prose.

IV. MAIN TEXT

The CCR house font style is Garamond size 11. Quotations of more than five lines/two sentences must be indented and in a smaller font size 10.

- Use UK English such as ‘s’ rather than ‘z’ spellings, eg recognise, nationalise.
- Numbers from one to ten are spelt out in words unless they refer to section or schedule numbers in statutes.
- Dates: 1 January 1999; the 1980s and 1990s (not 1990’s).
- Use per cent not % (e.g. eight per cent or 38 per cent).

1 **Subheading levels**

- Level one heading: bold, all capitalised, numbered I, II, III etc.
- Level two heading: bold, sentence case, numbered A, B, C etc.
- Level three heading: italic, sentence case, numbered 1, 2, 3 etc.
- Level four heading: plain, sentence case, numbered aa, bb, cc etc.
- Level five heading: italic, sentence case, numbered i, ii, iii etc.
For example:

I INTRUDUCTION

A Understanding what the Constitution requires

1 The meaning of the right to vote
   aa Democracy and the right to vote
      i South African cases

2. Case law

Depending on author preference, it is possible to either include the full case name (excluding the citation) in the main body of the text the first time it is referred to, for example Occupiers of 51 Olivia Road Berea Township and 197 Main Street Johannesburg v City of Johannesburg & Others (‘Olivia Road’), or to provide the full case name and citation in a footnote and refer only to the abbreviated case name in the main body of the text, for example Olivia Road.

3. Statutes

The full name of the statute including its number and year needs to appear in the main text. In other words, do not place the number and year of an Act in a footnote. This can be followed by an abbreviation in brackets, e.g. Promotion of Administrative Justice Act 3 of 2000 (PAJA), subsequent references are to ‘PAJA’.

Use ‘s’ or ‘ss’ (plural) instead of ‘section(s)’ unless it is the beginning of a sentence.

4. Quotations

Quotations should be clearly indicated by single quotation marks, with double quotation marks used for quotes within quotes. Where a quotation is two sentences long or runs to more than five lines, it must be in a smaller font, indented as a separate paragraph, with a line space above and below, and with no quotation marks or leader dots. Pay attention to accuracy when quoting directly.

5. Abbreviations

Abbreviations may be used for case names, eg (Olivia Road). However, the case name must be set out in full with full double-barreled citation the first time it is referred to, followed by an italicised abbreviation in brackets. If a case is abbreviated in a footnote (rather than in the main text), it is preferable for it to appear at the end of the sentence. The abbreviation can then be used throughout the main text and for cross referencing purposes in footnotes, eg the Court in Olivia Road held or the Olivia Road Court held.

• Abbreviated references to legislation can also be used in the body of the text (e.g. PAJA, PAIA or the Administrative Justice Act, the Information Act).
• Council for Conciliation, Mediation and Arbitration (CCMA) can occur within the text itself.
• The Constitutional Court is always abbreviated as ‘the Court’; all other courts are referred to as ‘the court’.
6. Avoid awkward or archaic turns of phrase

- Avoid polite legal clichés and wasted words such as ‘the learned judge’, ‘the learned author’, ‘with respect’, ‘with the greatest respect’, ‘it is submitted’ or ‘the authors humbly submit’.
- Judges can be referred to as Judge or Justice, preferably should be referred to as ‘Smith J’ or ‘Smith JA’ or ‘Smith LJ’. Avoid formulations such as ‘his Lordship’ or ‘the honourable’.
- Please eschew archaic uses of the first person. So, ‘it is my view’ or ‘I argue’ is preferable to ‘it is the view of the present author’, ‘it is this writer’s argument’. We know it’s your well-grounded belief.
- Write in the active voice. In short, ‘is’ can almost always be eliminated by a dynamic verb often found in adjectival form in the same sentence.

V. FOOTNOTES

The Journal makes use of footnotes, not parenthetical references. All articles, notes, comments, book reviews and contributions to the current developments section must make use of footnotes. Footnotes are in Garamond size 9.

1. Case law – South African cases

We expect double-barrelled citations for all South African cases. In most instances, this would mean SAFLII’s (ZACC, ZASCA or, if a High Court, something like ZAKZNHC) followed by Juta’s SALRs (e.g. 2008 (3) SA 208 (CC))

- Occupiers of 51 Olivia Road Berea Township and 197 Main Street Johannesburg v City of Johannesburg & Others [2008] ZACC 1, 2008 (3) SA 208 (CC) (‘Olivia Road’).

Case names in italic with ‘v’ for versus.

In the absence of an SALR citation, it is acceptable to use a Butterworths citation:


However, triple-barrel citations are fine.

- Occupiers of 51 Olivia Road Berea Township and 197 Main Street Johannesburg v City of Johannesburg & Others [2008] ZACC 1, 2008 (3) SA 208 (CC), 2008 (5) BCLR 475 (CC).

- Use commas rather than semi-colons between various citations (e.g. [2008] ZACC 1, 2008 (3) SA 208 (CC)).

Cross referencing cases

- Olivia Road (note 8 above) at para 45. (Where the case is not cited in the immediately preceding footnote).
- Use ‘Ibid.’ (Where the case and the paragraph reference (or the page reference in the case of a book or a journal) is the same as that in the immediately preceding footnote).
- Use ‘Ibid at para 45.’ (Where the case cited is identical, but the paragraph is not.)

Use paragraphs rather than page references wherever possible. This should always be possible for South African cases in roughly the last decade. All South African Constitutional Court decisions and most Supreme Court of Appeal, Land Claims Court and High Court decisions follow this practice.
2. Case law – foreign cases

Use US citations rather than S Ct or another citation form.
Again, double-barrelled citations are welcome.
Try to avoid: Romer v Evans 116 S Ct 1620, 1627 (1996). Please look up the US citation; it is always available online.
Avoid using abbreviated names of litigants, eg use Regents of the University of California not Regents of the Univ. of Cal.
Examples:

3. Bracketing

We encourage the use of brief parenthetical explanations of case holdings and quotations.
Examples:
- S v Makwanyane [1995] ZACC 3, 1995 (3) SA 391 (CC)(Court holds that death penalty constitutes a violation of rights to life and human dignity.)
- Wisconsin v Yoder 406 US 205, 123 SCt 456 (1972)(Supreme Court finds that compulsory school attendance for children of Amish religious community impairs right to free exercise of religion under 1st Amendment.)
- If you are quoting from a text, then provide both the page number and use quotation marks where appropriate:
  - E Mureinik ‘A Bridge to Where? Introducing the Interim Bill of Rights’ (1994) 10 South African Journal of Human Rights 31 (Mureinik argues that: ‘The drafters designed the Bill of Rights, and the Constitution as a whole, to foster a culture of law based upon justification, and no longer on mere authority and coercion.’)

Capitalise the first word inside a bracket (Court holds …) or (Dworkin contends …)
Do not use spaces in between brackets.
- S v Makwanyane [1995] ZACC 3, 1995 (3) SA 391 (CC)[No Space](Court holds that death penalty constitutes a violation of rights to life and human dignity.)

2. Journal articles

When citing journal articles give author’s initial and name, full title in quotation marks, year in parenthesis, volume number, full (not abbreviated) title of journal (italicised), first page of article, page referred to. Avoid the use of ‘at’ between first page and page referred; use a comma instead, e.g. 315, 325 (not 315 at 325).
The SAJHR should be cited as South African Journal on Human Rights. The Columbia LR should be cited as Columbia Law Review.
3. Books

When citing books, give author’s first initial and name, full title (italicised), edition, year, page reference. There is no need to state the place of publication and publisher. Page numbers should not be preceded by ‘p’ or ‘pp’.

Co-authors must be joined by an ampersand (&) rather than ‘and’.


Translations should be indicated thus: K Marx *Das Capital* (1867) (trans J Mander, 1976) 121.

4. Chapters in books

Author’s initial and name, full title in quotation marks, initial and name of editor(s), full title (italicised), year, first page of article, and specific page referred to in the text.


Subsequent references:

• Ibid at 9‒12.
• Cohen (note 1 above) at 9‒10.

5. Statutes

• Promotion of Administrative Justice Act 3 of 2000 (PAJA)
• Local Government: Municipal Finance Management Act 56 of 2003 (MFMA)
• Childrens Act 38 of 2005

Depending on author preference, it is possible to use the abbreviation for the name of an Act in subsequent references. If referring to a specific section/s, use the abbreviation ‘s’ or ‘ss’ except at the beginning of a sentence.

• PAJA s 6
• MFMA s 3
• Childrens Act s 1

6. The Constitution


The Interim Constitution requires a footnote. ‘The Interim Constitution has been repealed.’ Thereafter, ‘Constitution’ or ‘Final Constitution’ and ‘Interim Constitution’ may be used in the text and notes.

Subsequent references

• Constitution s 181(1)
• IC s 50
7. **Law Reform Commission papers**


8. **Parliamentary debates**

- NCOP Debates col 125 (24 February 1999)

9. **Treaties & international instruments**

Give ILM reference where available, failing which give UNTS reference or full UN Doc or OAU Doc reference.

- General Agreement on Tariffs and Trade, 30 Oct 1947, Protocol Amending the General Agreement to Introduce Part IV on Trade and Development and to Amend Annex I (8 Feb 1965) 572 *UNTS* 320.

For most *well-known multilateral treaties*, there’s no need for a bibliographical reference.

- Universal Declaration of Human Rights (1948)
- International Covenant on Civil and Political Rights (1966)

10. **Newspaper articles and Internet sources**

Newspaper articles and many internet sources can be treated much like other written sources: Author ‘Article Name’ *Newspaper/Internet Source* (Date), available at http://www.xxx. It is not necessary to record the date that the site was last visited.

VI. MORE GENERAL RULES

1. Eliminate ‘See’ from the beginning of all footnotes.

Exceptions:
We allow ‘see’ when they are buried in a footnote.

• For more on the contention that the idea of self-government should be considered to be the core component of a Constitution, even one with a justiciable Bill of Rights, see F Cachalia ‘Separation of Powers, Active Liberty and the Allocation of Public Resources: The E Tolling Case’ (2005) 132 South African Law Journal 285.

We allow ‘See also’ after primary citations.


2. Use ‘&’ rather than ‘and’ in footnotes


3. Titles are in first letter caps, no matter whether they are articles, books, newspaper articles or online documents.


• S Woolman A Politics of Accountability: How South Africa’s Judicial Recognition of the Binding Legal Effect of the Public Protector’s Recommendations Had a Catalysing Effect that Brought Down a President (2019).


4. On the use of ‘Section’, ‘section’ or ‘s’ with respect to the Constitution.

Write out ‘Section’ (capital S as in ‘Section 25 of the Constitution) to begin a sentence. Lower case ‘section’ is used in other instances (e.g. ‘In FNB, the Constitutional Court provided its first full length analysis of section 25 of the Constitution). However, it is not necessary to keep
repeating ‘section 25 of the Constitution’ throughout the text. The idea is to make it clear that we are talking about the Constitution, as opposed to another form of law. However, to repeat ‘section 25 of the Constitution’ each time would be truly unwieldy, cumbersome and just plain ol’ bad writing. In footnotes, the preferred form is Constitution s 25.

5. Please don’t forget to use ‘at’.
   • *AB CC* (note 4 above) at para 197.

6. Some footnotes are complicated.

We still try to make each component part as simple as style allows. Some citations warrant a brief description in a parenthetical; some require a number of sentences; some require only a prior reference such as “(note x above)” and a page number.

Example:

55 Gerrand (note 52 above) at 142 (Observes that ‘in terms of ancestral beliefs, family systems have rigid boundaries based on blood ties.’) Gerrand’s view is echoed by the KwaZulu-Natal Commissioner for Traditional Leadership Disputes and Crimes, Professor Jabulani Mphalala, who has stated that ‘it would take years before there was a flexibility of mind about adoption among most South Africans. […] Ancestral spirits look after their relatives and no-one else. In our religion, in our culture, this thing is ring-fenced.’ C Dardagan ‘Red-Tape Slowing down Adoptions’ *IOL* (21 February 2014), available at https://www.iol.co.za/lifestyle/family/parenting/red-tape-slowing-down-adoptions-1650829. See also Mokomane & Rochat (2010)(note 52 above) at ix; Rochat, Mokomane & Mitchell (note 52 above) at 124 (A study participants remarks: ‘When you are born, there are certain things that ancestors require of us. They know who our child is and where he is. Just imagine if you adopt a Biyela child and join the child to the Mthembus. There will be war between the Biyela and Mthembu ancestors, both ancestors will fight over who owns the child.’)