

Affirmative Action and Intensity of Review: *South African Police Service v Solidarity obo Barnard*

Chris McConnachie*

I INTRODUCTION

In assessing the merits of an affirmative action measure, as with the merits of any law or conduct, a court must engage in at least two levels of reasoning.¹ At the first level, a court must consider the reasons for and against the measure; its substantive merits. At the second level, a court must determine the appropriate intensity of review. This involves deciding whether, and to what extent, it is institutionally appropriate for a court to interrogate the substantive merits. In practice, these first and second levels of reasoning are intertwined. However, it is important to recognise that they are conceptually distinct parts of a court's reasoning process, even if they are inseparable in application.

This article is concerned with this second level of reasoning: the process of determining the appropriate intensity of review in affirmative action cases. The Constitutional Court's decision in *South African Police Service v Solidarity obo Barnard* was widely expected to provide guidance and clarity on this issue.² The majority judgment did not do so, but the divergent views expressed by the judges are instructive. In this article, I use *Barnard* to explore how the Court's approach to the intensity of review ought to develop in future affirmative action decisions.

By the time the Court handed down judgment in *Barnard*, more than ten years had passed since *Minister of Finance v Van Heerden*.³ *Van Heerden* was the Court's first and only decision scrutinising an affirmative action measure for compliance with

* Advocate, Thulamela Chambers, Johannesburg Bar; Research Associate, Rhodes University. I would like to thank the participants at the Constitutional Court Review workshop in November 2015 for their valuable input. Special thanks to David Bilchitz and Khomotso Moshikaro for their editorial input and guidance and to the anonymous reviewers for their helpful comments and suggestions. I am grateful to Tarunabh Khaitan and colleagues at Oxford for the many conversations that helped to shape my arguments in this article.

¹ See further E Kavanagh 'Deference or Defiance? The Limits of the Judicial Role in Constitutional Adjudication' in G Huscroft (ed) *Expounding the Constitution: Essays in Constitutional Theory* (2009) 190 ('Deference or Defiance?'); E Kavanagh 'Defending Deference in Public Law and Constitutional Theory' (2010) 126 *Law Quarterly Review* 222, 230; C Chan 'Proportionality and Invariable Baseline Intensity of Review' (2013) 33 *Legal Studies* 1, 12 ('Baseline Intensity'). Both Kavanagh and Chan draw on J Raz *Practical Reason and Norms* (2nd Edition, 1999).

² *South African Police Service v Solidarity obo Barnard* [2014] ZACC 23, 2014 (6) SA 123 (CC), 2014 (10) BCLR 1195 (CC) ('*Barnard*').

³ *Minister of Finance & Another v Van Heerden* [2004] ZACC 3, 2004 (6) SA 121 (CC), 2004 (11) BCLR 1125 (CC) ('*Van Heerden*').

s 9(2) of the Constitution.⁴ That decision transformed s 9(2) from an interpretive guide into a fully fledged test for adjudicating the validity of affirmative action measures.⁵ The resulting ‘*Van Heerden* test’ was the Court’s first attempt to sketch the appropriate intensity of review in affirmative action cases.

In the decade after *Van Heerden*, lower courts largely ignored the Constitutional Court’s test.⁶ That much was evident as *Barnard* worked its way through the court system over a period of more than seven years. Three separate courts, including the Supreme Court of Appeal, failed to apply the *Van Heerden* test.

Given its path to the Constitutional Court, *Barnard* was tipped to be the Court’s moment to reassert and refine the *Van Heerden* test. The Court was also expected to clarify how the *Van Heerden* test should be tailored to the employment context and the specific provisions of the Employment Equity Act (EEA).⁷ In doing so, it was hoped that the Court would provide further guidance on the appropriate intensity of review.

The majority judgment and three separate concurrences in *Barnard* failed to live up to these expectations. Moseneke ACJ, writing for the majority dodged the central issues by holding that the unfair discrimination challenge was not properly before the Court. In passing, he affirmed that the *Van Heerden* test applies in assessing the validity of affirmative action measures under the EEA. However, he suggested, in obiter, that a different test is required in assessing the implementation of these measures, with rationality as its core. In his concurring judgment, Jafta J endorsed this rationality standard. The concurring judgment of Cameron, Froneman JJ and Majiedt AJ (Cameron et al) grappled with the unfair discrimination challenge. However, they did so by developing a new test for the validity of affirmative action under the EEA, one based on fairness. In a separate concurrence, Van der Westhuizen J was the only member of the court to apply the *Van Heerden* test in assessing the implementation of the affirmative action measure. The end result was a host of different approaches to the intensity of review with little guidance for future decisions.

In a detailed note on *Barnard*, Cathi Albertyn argues that the Court missed the opportunity to develop its affirmative action jurisprudence.⁸ In this article, I build on Albertyn’s analysis of the missed opportunities, focusing specifically on the intensity of review in applying the *Van Heerden* test.

⁴ Constitution of the Republic of South Africa, 1996. Section 9(2) states that: ‘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.’

⁵ The Constitutional Court is averse to using the term ‘affirmative action’, preferring the terms ‘remedial’ or ‘restitutionary’ measures (see *Van Heerden* (note 3 above) at para 29). However, I will refer to affirmative action as it is the term with the widest currency.

⁶ This has been a common pattern in the Labour Court and Labour Appeal Court. See further A Rycroft ‘Transformative Failure: The Adjudication of Affirmative Action Appointment Disputes’ in O Dupper & C Garbers (eds) *Equality in the Workplace: Reflections from South Africa and Beyond* (2009).

⁷ Act 55 of 1998.

⁸ C Albertyn ‘Adjudicating Affirmative Action Within a Normative Framework of Substantive Equality and the Employment Equity Act – An Opportunity Missed?’ *South African Police Service v Solidarity obo Barnard* (2015) 132 *South African Law Journal* 711.

I argue that the *Barnard* Court should have taken the opportunity to clarify the intensity of review in three respects. First, in support of Albertyn and other commentators, I argue that the Court should have affirmed that the *Van Heerden* test is applicable to affirmative action measures and their implementation under the EEA. Second, I also support the view that the Court should have clarified that the standard of proof embodied in the *Van Heerden* test is a proportionality analysis. Third, and most significantly, I go beyond the existing commentaries by exploring how and why the *Van Heerden* test may be applied with variable intensity.

Existing contributions to the South African literature focus on the standard of review to be applied in affirmative action cases.⁹ However, there has not yet been any serious engagement with the ways in which the chosen standard of review may be applied with different intensity depending on the context. As I will demonstrate, *Barnard* illustrates how this variable intensity of review will often be decisive. Given that so much turns on this variable intensity of review, it is an area of the Court's jurisprudence where far greater guidance and transparency is needed. In addition, this variable intensity of review is an important tool. It allows the Court to increase the intensity of review where this is necessary to prevent abuses and to come to the assistance of historically disadvantaged groups. I contend that the Court must openly justify its chosen intensity of review by reference to a set of three principles: the interests at stake, relative institutional competence, and considerations of democratic legitimacy. This is one of the central tasks for the Court to address in future cases.

I will develop this argument in five parts. Part II explains the nature and importance of the intensity of review in affirmative action cases. In Part III, I discuss the *Van Heerden* test and highlight where the test requires clarification and development. In Part IV, I provide some background to the Constitutional Court's decision in *Barnard*, setting out the facts and the lower courts' decisions. In Part V, I explore the Constitutional Court's reasoning in *Barnard* by assessing the majority and concurring judgments. I focus on the different standards of review adopted by the Court and the intensity with which these standards were applied. Finally, in Part VI, I set how the Court should have developed the *Van Heerden* test, emphasising the need for a principled approach to the variable intensity of review in future cases.

⁹ See, eg, C Albertyn & B Goldblatt 'Equality' in S Woolman & M Bishop (eds) *Constitutional Law of South Africa* (2nd Edition, 2008) 35-33-35-39; JL Pretorius 'R v Kapp: A Model for South African Affirmative Action Jurisprudence?' (2009) 126 *South African Law Journal* 398 ('R v Kapp'); JL Pretorius 'Fairness in Transformation: A Critique of the Constitutional Court's Affirmative Action Jurisprudence' (2010) 26 *South African Journal on Human Rights* 536 ('Fairness in Transformation'); JL Pretorius 'Accountability, Contextualisation and the Standard of Judicial Review of Affirmative Action: *Solidarity obo Barnard v South African Police Services*' (2013) 130 *South African Law Journal* 31 ('Standard of Review'); JL Pretorius 'The Unresolved Search for the Proper Standard of Review of Affirmative Action: *Solidarity obo Barnard v SAPS*' (2013) 38 *Journal of Juridical Sciences* 128 ('Unresolved Search'); Albertyn *ibid*.

II INTENSITY OF REVIEW

A Importance and Substance

Before going further, it is necessary to say something about the importance of studying the intensity of review in affirmative action cases. On first glance, second-level reasoning about the appropriate intensity of review can appear formalistic and value-free. First-level reasoning about the substantive merits of affirmative action measures can often appear more urgent and important. First-level reasoning directly engages subjects such as the aims and values that should inform affirmative action measures, the effectiveness of these measures, and the balance between the benefits of specific measures and their costs.¹⁰ However, this does not mean that the intensity of review is any less significant or value-driven.

The United States Supreme Court's problematic affirmative action jurisprudence is a cure for any doubts about the importance and value-laden nature of the intensity of review. The Supreme Court's 'strict scrutiny' approach to race-based affirmative action has had profound effects, forcing most race-based affirmative action into hiding or retreat.¹¹ This standard of review sets a high bar by requiring any race-based measure to serve a compelling purpose and to be the only means of achieving that purpose. The ongoing debates on the Supreme Court over strict scrutiny also show that the choice of intensity of review is driven by divergent values.¹² The liberal and conservative judges' views are based on fundamentally different assumptions about racial injustice, the importance of affirmative action, and the role of the courts in evaluating these measures.

Fortunately, the South African Constitutional Court is not heading in the direction of a restrictive, US-style jurisprudence.¹³ The Court has rightly distanced itself from the strict scrutiny approach, holding it up as a benchmark of failure.¹⁴ Nevertheless, the challenge of finding an appropriate judicial role in evaluating affirmative action in South Africa is no less important.

While our courts are not chafing against the restraints of an overly restrictive affirmative action jurisprudence, they are faced with the opposite problem. They currently lack appropriate guidance on the appropriate intensity of review in these cases. *Barnard* and the string of other affirmative action cases currently before the

¹⁰ These issues have been canvassed extensively in the South African literature. See, eg, O Dupper 'In Defence of Affirmative Action in South Africa' (2004) 121 *South African Law Journal* 187; O Dupper 'Affirmative Action: Who, How and How Long?' (2008) 24 *South African Journal on Human Rights* 425; P de Vos 'The Past is Unpredictable: Race, Redress and Remembrance in the South African Constitution' (2012) 129 *South African Law Journal* 73.

¹¹ For an overview of this jurisprudence, see S Fredman *Discrimination Law* (2nd Edition, 2011) 118–125; S Fredman *Comparative Study of Anti-Discrimination and Equality Laws of the US, Canada, South Africa and India* (2012) 64–65; R Siegel 'Equality Divided' (2013) 127 *Harvard Law Review* 1.

¹² See, eg, the debates between the majority and minority in *Adarand Constructors Inc v Peña* 515 US 200 (1995); *Grutter v Bollinger* 539 US 306 (2003); *Fisher v University of Texas* 570 US – (2013); *Schnette v Coalition to Defend Affirmative Action, Integration and Immigration Rights and Fight for Equality By Any Means Necessary* 572 US – (2014).

¹³ For a comparison of the South African and US approaches, see K Naff & O Dupper 'Footprints Through the Courts: Comparing Judicial Responses to Affirmative Action Litigation in South Africa and the United States' (2009) 25 *International Journal of Comparative Labour Law and Industrial Relations* 157.

¹⁴ See *Van Heerden* (note 3 above) at paras 29, 147.

courts show that this uncertainty is fertile ground for litigation, particularly for groups seeking to frustrate affirmative action measures.¹⁵

There is no escaping the fact that the task of developing an appropriate intensity of review is a complex, technical exercise. That does not make it formalistic, in the sense of being unmoored from values and deeper debates about the merits of affirmative action.¹⁶ However, it remains a task laden with complexity. Part of the complexity is conceptual, as the intensity of review has many moving parts. In what follows, I briefly explain the intensity of review, its connection with deference and the ways in which these concepts manifest in judicial decisions.

B Intensity of Review, Deference and Their Manifestations

The intensity of review broadly refers to the strictness with which a court assesses the validity of laws and actions. This is generally related to the degree of deference that a court shows to the decision-maker in each case. Timothy Endicott helpfully defines deference as a court's willingness to 'leave the answer to some question, to some extent, to the initial decision-maker'.¹⁷ It is also common to see deference described as the process of giving greater 'weight' to the decision-maker's reasons for a law or action than those reasons would otherwise deserve.¹⁸ Both descriptions capture the idea that deference involves courts suspending judgment on the substantive merits of a law or action, at least to some degree.

A court's intensity of review and its degree of deference are generally inversely related. The more deferent the court the less intensely it will scrutinise laws and conduct, and vice versa. However, this relationship does not always hold. At times a court may go so far as to supply its own arguments and evidence in favour of a law or action, going beyond what is presented by the parties. In doing so, the court is no longer deferring to the decision-maker, in the sense of leaving certain matters unquestioned or giving greater weight to reasons provided. Instead, it is assuming a more active role in defending the law or action. This is not necessarily wrong, but it is important to watch for slippage where the language of deference is used to mask a more proactive or even partisan approach.

A court can adjust the intensity of review in two primary ways: by using different *standards of review* and by varying the intensity with it *applies* these standards.¹⁹

¹⁵ See Albertyn (note 8 above) 712–713; S Budlender, G Marcus & N Ferreira *Public Interest Litigation in South Africa: Strategies, Tactics and Lessons* (2014) 16–17 (on the attempts by the trade union Solidarity to resist affirmative action).

¹⁶ On the distinction between formalism and substantive reasoning, see PS Atiyah & RS Summers *Form and Substance in Anglo-American Law: A Comparative Study of Legal Reasoning, Legal Theory, and Legal Institutions* (1987) 1, 2, 5.

¹⁷ T Endicott *Administrative Law* (2nd Edition, 2011) 225.

¹⁸ Kavanagh 'Deference or Defiance?' (note 1 above) at 185; C Chan 'Deference, Expertise and Information-Gathering Powers' (2013) 33 *Legal Studies* 598, 600 ('Expertise').

¹⁹ For further discussion of this distinction see J Rivers 'Proportionality and Variable Intensity of Review' (2006) 65 *Cambridge Law Journal* 174 ('Variable Intensity'); Chan 'Baseline Intensity' (note 1 above) 5. Some have questioned the value of distinguishing between different standards of review, suggesting that it would be simpler to ask whether there are compelling reasons for a law or action in each context. That debate is beyond the scope of this article, but it suffices to say that the different standards of review have the benefit of providing analytical guidance, as opposed to a free-wheeling injunction to analyse the strength of the reasons.

Standards of review are the different sets of questions that a court asks in reviewing a law or action.²⁰ These standards of review exist on a spectrum, ranging from higher to lower levels of intensity. At the more intense end of this spectrum are standards of review such as the US Supreme Court's strict scrutiny approach, requiring a compelling purpose for race-based affirmative action measures and for these measures to be the only means of achieving this purpose.²¹ On the lower end of the spectrum are standards of review like rationality, which assesses whether the law or action is rationally connected to some legitimate purpose.²² Rationality is now the Canadian Supreme Court's favoured standard of review in assessing affirmative action measures, as set out in its 2008 decision in *R v Kapp*.²³ Proportionality and its more loosely defined cousin, reasonableness,²⁴ occupy a space in between strict scrutiny and rationality, involving a balancing of the benefits and harms of laws and actions.²⁵

Choosing a standard of review is not the end of the matter. Each of the standards of review can be applied with variable intensity from case to case. As Julian Rivers points out, merely talking about different standards of review fails to capture this variability.²⁶ For example, a court can subtly adjust the proportionality analysis in many ways, placing a thumb on the scales.²⁷ Similarly, the rationality analysis can also be applied more or less stringently, as commentators on the Court's rationality jurisprudence have repeatedly noted.²⁸

The intensity of review in applying these standards is intimately linked with the burden of proof and a court's willingness to play a proactive role in uncovering

²⁰ See generally S Seedorf & S Sibanda 'Separation of Powers' in S Woolman & M Bishop (eds) *Constitutional Law of South Africa* (2nd Edition, 2008) 12-59ff (OS 06-08).

²¹ See further Fredman (note 11 above) 118-125.

²² See, eg, *Prinsloo v Van der Linde and Another* [1997] ZACC 5, 1997 (3) SA 1012 (CC) at para 26 (differentiation must be rationally connected to a legitimate government purpose). Rationality has acquired other dimensions in the Court's s 1(c) rationality jurisprudence.

²³ *R v Kapp* [2008] 2 SCR 483 ('*Kapp*') (Canadian Supreme Court dismissed a challenge to a law allowing First Nations people to be given the exclusive licence to fish for salmon at valuable fishing grounds for a 24-hour period).

²⁴ On the relationship between proportionality and reasonableness, see, eg, C Steinberg 'Can Reasonableness Protect the Poor? A Review of South Africa's Socio-Economic Rights Jurisprudence' (2006) 123 *South African Law Journal* 264, 278-280; C Hoexter *Administrative Law in South Africa* (2nd Edition, 2013) 343-350 ('*Administrative Law*').

²⁵ Reflected in the Court's account of proportionality in *S v Manamela and Another (Director-General of Justice Intervening)* [2000] ZACC 5, 2000 (3) SA 1 (CC), 2000 (5) BCLR 491 (CC) at para 32 ('Court must engage in a balancing exercise and arrive at a global judgment on proportionality and not adhere mechanically to a sequential check-list').

²⁶ Rivers (note 19 above) 202ff.

²⁷ See, eg, the debate between the majority and minority in *Prince v President of the Law Society of the Cape of Good Hope and Others* [2002] ZACC 1, 2002 (2) SA 794 (CC), 2002 (3) BCLR 231 (CC) (Disagreement over the suitability and necessity of criminalising marijuana without exception for religious use).

²⁸ See C Hoexter 'Just Administrative Action' in I Currie & J De Waal (eds) *The Bill of Rights Handbook* (6th Edition, 2013) 688ff; M Bishop 'Rationality is Dead! Long Live Rationality! Saving Rational Basis Review' in S Woolman & D Bilchitz (eds) *Is This Seat Taken? Conversations at the Bar, the Bench & the Academy* (2012); A Price 'Rationality Review of Legislation and Executive Decisions: Poverty Alleviation Network and *Albuti*' (2010) 127 *South African Law Journal* 580; M du Plessis & S Scott 'The Variable Standard of Rationality Review: Suggestions for Improved Legality Jurisprudence' (2013) 130 *South African Law Journal* 597; L Kohn 'The Burgeoning Constitutional Requirement of Rationality and the Separation of Powers: Has Rationality Review Gone Too Far?' (2013) 130 *South African Law Journal* 810.

evidence and argument. Deference to the state is often shown in the way that courts give the state the benefit of the doubt. Judicial notice – the judicial recognition of facts that are considered to be general knowledge or are easily ascertainable from sources of incontrovertible authority²⁹ – also plays a role here. Courts may invoke judicial notice to fill gaps in the decision-maker’s evidence. The court may also raise arguments or issues not addressed by the decision-maker as a way to bolster the reasons for or against a law or action.

In the existing South African literature on affirmative action, commentators have focused almost exclusively on the appropriate standards of review.³⁰ There has not yet been any sustained analysis of how and why these standards of review may be varied depending on the context. As I will argue in Parts V and VI, this is a crucial area for further development of the Court’s affirmative action jurisprudence.

III THE *VAN HEERDEN* TEST

The Constitutional Court’s decision in *Van Heerden* was its first opportunity to grapple with the appropriate intensity of review in affirmative action cases. Before *Van Heerden*, the Court approached affirmative action measures through the lens of the prohibition of unfair discrimination under s 9(3) of the Constitution.³¹ The validity of these measures was tested under the ‘*Harksen* test’ for unfair discrimination,³² developed in *Harksen v Lane* NO.³³

²⁹ PJ Schwikkard & SE van der Merwe ‘Judicial Notice’ in PJ Schwikkard & others (eds) *Principles of Evidence* (4th Edition, 2015) ch 27.

³⁰ See note 9 above.

³¹ Section 9 of the Constitution, the equality clause, provides that:

- (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.’

The Court has not yet decided an unfair discrimination challenge based on s 9(4) of the Constitution.

³² See, eg, *President of the Republic of South Africa and Another v Hugo* [1997] ZACC 4, 1997 (4) SA 1 (CC), 1997 (6) BCLR 708 (CC) (Court found that the President’s decision to pardon mothers of young children was not unfairly discriminatory); *City Council of Pretoria v Walker* [1998] ZACC 1, 1998 (2) SA 363 (CC), 1998 (3) BCLR 257 (CC) (Court found that a municipality’s debt collection policy unfairly discriminated against white residents). It is arguable that in both cases, the measures would not have passed the *Van Heerden* test and would still have needed to be considered under the *Harksen* test. See further Albertyn & Goldblatt (note 9 above) 32–35.

³³ *Harksen v Lane NO and Others* [1997] ZACC 12, 1998 (1) SA 300 (CC), 1997 (11) BCLR 1489 (CC) at para 53(‘*Harksen*’) (Court synthesised principles from earlier case law in setting out a structured test for unfair discrimination).

Section 9(2) of the Constitution merely served as an interpretive guide to this analysis.³⁴

Van Heerden changed this by stipulating that affirmative action measures should first be considered for compliance with s 9(2). If a measure satisfies s 9(2) then it is immunised from further scrutiny under s 9(3).³⁵ Only if it fails the *Van Heerden* test should it then be subjected to the *Harksen* test.

In creating the *Van Heerden* test, the Court was reacting to certain features of the *Harksen* test which required inappropriate forms of scrutiny for affirmative action. As a result, it is important to understand the *Harksen* test before considering the *Van Heerden* test in greater depth.

A From *Harksen* to *Van Heerden*

The *Harksen* test for unfair discrimination has two stages. First, a court must determine whether there is discrimination, which involves the imposition of burdens or the withholding of benefits on grounds listed in s 9(3) of the Constitution, or grounds analogous to these listed grounds. Second, if there is discrimination, then it must be determined whether the discrimination is unfair.

The unfairness analysis in the *Harksen* test is a complex and under-analysed part of the Court's jurisprudence. What is clear is that it involves two sets of enquiries: a) an analysis of the impact of the discrimination on the disfavoured group and b) an analysis of the justification for the discrimination, taking into account its impact.³⁶ This unfairness analysis is generally applied with varying intensity of review, as I have analysed in previous work.³⁷ The Court's standard of review in assessing the justification for discrimination often fluctuates between a rationality analysis and a proportionality analysis. There is also a great deal of variability in the intensity with which these standards of review are applied.³⁸

The *Harksen* test also includes a built in presumption of unfairness. Section 9(5) of the Constitution requires that, where discrimination has occurred on grounds listed in s 9(3), this discrimination must be presumed to be unfair, placing the

³⁴ See *National Coalition for Gay and Lesbian Equality and Another v Minister of Justice and Others* [1998] ZACC 15, 1999 (1) SA 6 (CC), 1998 (12) BCLR 1517 (CC) at paras 61–62 (explaining that s 9(2) requires 'substantive and remedial equality').

³⁵ *Van Heerden* (note 3 above).

³⁶ See further S Woolman & H Botha 'Limitations' in S Woolman & M Bishop (eds) *Constitutional Law of South Africa* (2nd Edition, 2008) 34-34–34-35 (OS 07-06); R Krüger 'Equality and Unfair Discrimination: Refining the *Harksen* Test' (2011) 128 *South African Law Journal* 479, 504–505; C McConnachie *What Is Unfair Discrimination? A Study of the South African Constitutional Court's Unfair Discrimination Jurisprudence* (DPhil Thesis, University of Oxford, 2014) ch 5.

³⁷ See C McConnachie 'Transformative Unfair Discrimination Jurisprudence: The Need for a Baseline Intensity of Review' (2015) 31 *South African Journal on Human Rights* 504, 516–524.

³⁸ *Ibid.*

burden squarely on the person alleged to be engaging in discrimination to prove otherwise.³⁹

The Court's efforts in *Van Heerden* to create a new test for assessing the validity of affirmative action should be seen against this backdrop.

B The Structure of the *Van Heerden* Test

Van Heerden concerned a challenge to a pension scheme for members of Parliament. The scheme paid bigger pension contributions to MPs who were elected after 1994. Those who served before 1994 were given a smaller contribution, reflecting the fact that the majority received generous pension packages under apartheid. The aim of this policy was to allow black MPs to develop pension savings comparable to their longer-serving, predominantly white colleagues. Frederick van Heerden, a white, National Party MP, challenged the scheme, arguing that it was unfairly discriminatory on the basis of race. The Court dismissed this challenge.

The key question facing the Court was how to go about evaluating this pension scheme. One option was to assess it under the s 9(3) prohibition of unfair discrimination, using the s 9(2) affirmative action provision merely as an interpretive aid. Another was to interpret s 9(2) as a standalone test for valid affirmative action measures, thus insulating valid affirmative action from further scrutiny under s 9(3) if it satisfies that test.⁴⁰

Moseneke J, writing for the majority, chose the latter option by interpreting s 9(2) as a standalone test. He held that affirmative action measures will only be scrutinised under the *Harksen* test if they fail the s 9(2) test.⁴¹ Moseneke J proceeded to outline three requirements for the validity of affirmative action measures, based on the wording of s 9(2).

First, the measure must '[target] persons or categories of persons who have been disadvantaged by unfair discrimination', requiring that members of historically disadvantaged groups should make up an 'overwhelming majority' of the beneficiaries of the measure.⁴²

Second, the measure must be 'designed to protect or advance such persons or categories of persons', meaning that it must be conducted for the purpose of benefitting disadvantage groups, it must not be 'arbitrary, capricious or

³⁹ In theory, this ought to require more stringent scrutiny in applying this analysis, although the Court has often wilfully ignored the s 9(5) presumption in practice. See, eg. *S v Jordan and Others* [2002] ZACC 22, 2002 (6) SA 642 (CC), 2002 (11) BCLR 1117 (CC) (Court found that the criminalisation of sex workers, but not their clients, is fair discrimination, despite the state offering no justification for the discrimination); *Volks NO v Robinson and Others* [2005] ZACC 2, 2005 (5) BCLR 446 (CC) (Court found that the exclusion of life partners from the Maintenance of Surviving Spouses Act 27 of 1990 was fair discrimination, despite the state and the executor conceding that the law was unfairly discriminatory).

⁴⁰ Pretorius 'R v Kapp' (note 9 above) terms these the 'integrative' and 'exemptive' approaches.

⁴¹ *Van Heerden* (note 3 above) at para 36.

⁴² *Ibid* at paras 38, 40. This accommodates cases of indirect affirmative action, where measures are not explicitly targeted at particular disadvantaged groups, but have the purpose and effect of benefitting disadvantaged groups nonetheless. On the nature and merits of indirect affirmative action, see T Khaitan *A Theory of Discrimination Law* (2015) ch 8.

display naked preference’, and it must be ‘reasonably likely’ to protect or advance historically disadvantaged people.⁴³

Third, the measures must ‘promote the achievement of equality’.⁴⁴ Moseneke J suggested that this involves some consideration of the impact of the affirmative action measure on those who are excluded:

Determining whether a measure will in the long run promote the achievement of equality requires an appreciation of the effect of the measure in the context of our broader society. It must be accepted that the achievement of this goal may often come at a price for those who were previously advantaged. ... However ... a measure should not constitute an abuse of power or impose such substantial and undue harm on those excluded from its benefits that our long-term constitutional goal would be threatened.⁴⁵

In his commentary on the case, JL Pretorius has argued that the *Van Heerden* test verges on a rationality analysis.⁴⁶ That would be true if the Court had left the test at the first two requirements. However, the inclusion of the third requirement of the test clearly contemplates some form of proportionality analysis, which involves weighing up the benefits of the affirmative action measure against its impact on those who are excluded.⁴⁷ A court could only determine whether the harm to the excluded is sufficiently ‘substantial’ and ‘undue’ by engaging in an exercise of proportionate balancing.

C The Justification for a Separate Test

The *Van Heerden* test is not entirely distinct from the *Harksen* test. Both tests involve some balancing of interests and consideration of the impact of the measure. However, there are two important differences between these tests. These differences are motivated by problems in applying the *Harksen* test to affirmative action in the first instance.

The first problem in applying the *Harksen* test to affirmative action is the expressive harm of placing concern for privileged groups ahead of the needs of historically disadvantaged groups.⁴⁸ Under the *Harksen* test, a Court is required to focus on the impact of the discrimination on the disfavoured individual or group.⁴⁹ Given that privileged groups are the most likely to be excluded from affirmative action measures, the *Harksen* test would place their interests front-and-centre in the analysis. This risks suggesting that the benefits of an affirmative action measure for historically disadvantaged groups are only of secondary concern. The third requirement of the *Van Heerden* test still leaves room for an assessment of

⁴³ *Van Heerden* (note 3 above) at para 41–43.

⁴⁴ *Ibid* at para 44–45.

⁴⁵ *Ibid* at para 44.

⁴⁶ See, eg, Pretorius ‘Fairness in Transformation’ (note 9 above) 537, 561–567.

⁴⁷ See J Brickhill ‘Testing Affirmative Action under the Constitution and the Equality Act’ (2006) *Industrial Law Journal* 2004, 2013; P de Vos (note 10 above) 93; Albertyn (note 8 above) 730–731.

⁴⁸ On the nature and significance of expressive concerns, see MD Adler ‘Expressive Theories of Law: A Skeptical Overview’ (2000) 148 *University of Pennsylvania Law Review* 1363, 1428–1447; T Khaitan ‘Dignity as an Expressive Norm: Neither Vacuous Nor a Panacea’ (2012) 32 *Oxford Journal of Legal Studies* 1, 5–9.

⁴⁹ *Van Heerden* (note 3 above) at para 80 (Mokgoro J noted this expressive problem).

the impact of affirmative action on disfavoured individuals and groups. However, it does not make this the focal point of the analysis, as the *Harksen* test does.

As a result, the *Harksen* and *Van Heerden* tests are different in the ordering of the impact analysis. The *Van Heerden* test first considers the benefits for the historically disadvantaged group and only then assesses its impact on those who are excluded or adversely affected. That ordering of the analysis may not have any practical significance, in the sense that this is unlikely to affect the outcome of decisions. However, it certainly has symbolic significance.

The second, and most significant, reason for this separate test is that the *Harksen* test would require courts to treat all affirmative action measures as presumptively unfair. That flows from the s 9(5) presumption of unfairness where discrimination occurs on grounds listed in s 9(3). Moseneke J held that it is inconsistent with the constitutional scheme to apply the s 9(5) presumption of unfairness in this way:

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. Secondly, such 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.⁵⁰

This passage points both to expressive and institutional difficulties in holding affirmative action to be presumptively unfair.

The presumption is expressively problematic as it sends the message that all affirmative action is wrongful unless proved otherwise, rather than being a constitutionally sanctioned means of addressing patterns of group disadvantage. That is clearly inconsistent with the scheme of s 9 and the values that animate it.

The presumption of unfairness is also institutionally harmful. As Moseneke J acknowledges, this presumption may set the intensity of review too high. A presumption of unfairness may make courts all too willing to second-guess the state in complex matters, emboldening conservative litigants and judges to attempt to curtail affirmative action wherever possible. A further reason for rejecting this presumption is that it would require courts to err on the side of invalidating affirmative action measures where there is doubt about where the balance of probabilities lies. A presumption is, in essence, a technical device to allocate the risk of uncertainty.⁵¹ In the case of affirmative action, the potential for uncertainty is high. The effects of affirmative action, both positive and negative, will generally be apparent only after the measure has been in force over many years. In some cases, these effects may never be capable of accurate quantification. As Moseneke J acknowledged, the ‘future is hard to predict’.⁵² As a result, a presumption of unfairness would unduly curtail affirmative action, as it

⁵⁰ Ibid at para 33.

⁵¹ For more detailed analysis of presumptions in human rights litigation, see J Rivers ‘The Presumption of Proportionality’ (2014) 7 *The Modern Law Review* 409, 415–418 (‘Presumption of Proportionality’). See also C Chan, ‘The Burden of Proof under the Human Rights Act’ (2014) 19 *Judicial Review* 46 (‘Burden of Proof’).

⁵² *Van Heerden* (note 3 above) at para 41.

may often be impossible for the state to muster the necessary evidence at the time of litigation to prove that the benefits of particular affirmative action measures will outweigh the costs, on a balance of probabilities.

The Court undoubtedly had a further institutional worry in mind, although this is not openly articulated in *Van Heerden*. The presumption of unfairness raises the spectre that otherwise good and justifiable affirmative action programmes may be invalidated due to the state's failure to mount a proper defence in court. Even in cases where ample evidence and argument are available, the state has a track record of offering poor defences, or no defence at all. In such cases, the presumption of unfairness would not require the Court to 'second-guess' the other branches of state, as that presumes that there is reasoning before the Court to be scrutinised. Instead, the presumption would prevent the Court from coming to the defence of affirmative action measures by going beyond the meagre case presented by the state.

In his commentary on affirmative action decisions, Pretorius has suggested that the presumption of unfairness ought to be applied to affirmative action.⁵³ His argument appears to be that this presumption is a necessary component of a culture of accountability.⁵⁴ However, the need for accountability does not automatically translate into an argument for presumptive unfairness. Accountability requires that the state should, at a minimum, give reasons for its actions. But a duty of reason-giving is separate from the question of whether the state should be required to put up sufficient evidence and argument to overcome a presumption. A presumption sets a default position in the case of uncertainty, providing that affirmative action is unlawful unless proved otherwise. The symbolic and institutional difficulties of regarding all affirmative action as being unlawful by default are sufficient reason to reject the presumption in these cases.

The absence of a presumption of unfairness is the most significant difference between the *Van Heerden* and *Harksen* tests. The result is that the burden of proof will generally fall on the complainant to show that the affirmative action measure fails to comply with the s 9(2) requirements. However, the absence of a presumption of unfairness does not prevent a court from shifting the burden of proof to the opposing party in appropriate cases, either by tinkering with the burden of non-persuasion or the evidential and argumentative burden.⁵⁵ I will explore this important dimension of the variable intensity of review in greater detail in Part VI.

D Tasks for Future Decisions

The *Van Heerden* test left much work to do. Three tasks, in particular, were set for future decisions.

First, clarity was needed on whether the *Van Heerden* test applies to affirmative action measures under the statutory equality instruments, including the EEA and the Promotion of Equality and Prevention of Unfair Discrimination Act

⁵³ Pretorius 'Standard of Review' (note 9 above) 37–42.

⁵⁴ Ibid 39–40.

⁵⁵ I explain this distinction at text to notes 145–147 below.

(Equality Act).⁵⁶ Both statutes mirror s 9 of the Constitution, as they contain prohibitions on unfair discrimination, with the qualification that affirmative action is permissible.

Second, the precise standard of review embodied in the *Van Heerden* test also required development and clarification. All three requirements of the test were open-textured and required further refinement.

Third, the Court also needed to establish precisely how and why this standard of review may be applied with variable intensity. This variable intensity of review is an important and necessary component of the *Van Heerden* test. While affirmative action should not be treated as inherently suspect, there are likely to be cases where affirmative action measures may aggravate the patterns of disadvantage experienced by historically disadvantaged groups. For example, an affirmative action measure in the workplace that systematically excluded women would need to attract a heightened level of scrutiny. The variable intensity of review will allow the courts to respond appropriately in these situations. However, *Van Heerden* provided no guidance as to when and how this intensity of review may be adjusted.

It is no criticism of the *Van Heerden* Court to point to these loose ends. The Court made an admirable first step in developing a test while leaving room for development and invention. It was perhaps too much to expect the Court to offer a fully formed, precise test for such a contested issue in its first genuine affirmative action decision. The clear intention was that the test would gradually develop through use and exposure to real-world problems. However, a full decade would pass before the Court would have another opportunity to revisit the scope and content of the *Van Heerden* test.

IV BACKGROUND TO *BARNARD*

A The Facts

The facts in *Barnard* will now be familiar to most. Renate Barnard, a white woman, was a captain in the South African Police Service (SAPS). On two occasions, she applied for promotion to vacant positions. She was twice rejected despite being judged the best candidate. On both occasions, the interview panel recommended her for appointment. In her first application, her Divisional Commander declined to support her appointment as this would not advance racial representivity.⁵⁷ In the second application, it was the National Commissioner who declined her appointment, despite the Divisional Commander's recommendation.⁵⁸ The post was subsequently re-advertised and then scrapped in a process of restructuring.⁵⁹

The National Commissioner's reasons for rejecting Barnard's promotion were set out in a brief letter.⁶⁰ The National Commissioner reasoned that Barnard's appointment would not advance employment equity at the relevant salary level.

⁵⁶ Act 4 of 2000.

⁵⁷ *Barnard* (note 2 above) at para 9.

⁵⁸ *Ibid* at para 14.

⁵⁹ *Ibid*.

⁶⁰ *Ibid* at paras 15–16.

He also emphasised that the post was not critical and service delivery would not be affected if the post was left vacant until a suitable candidate could be found.

In reaching his decision, the National Commissioner was implementing the SAPS employment equity plan. The SAPS is a 'designated employer' for the purposes of the EEA.⁶¹ As a result, it is legally required to have an employment equity plan setting targets for the advancement of members of 'designated groups', defined under the Act as black people, women, and people with disabilities.⁶² The SAPS employment equity plan sets detailed numerical targets for the representation of members of designated groups at different salary levels. Barnard had applied for appointment at salary level 9, a level where white men and women were overrepresented.⁶³

The Employment Equity Act sets out requirements for how employment equity measures should be implemented. Section 15(1) provides that these measures should be 'designed to ensure that suitably qualified people from designated groups have equal employment opportunities and are equitably represented in all occupational levels in the workforce of a designated employer'. Section 15(3) further provides that while preferential treatment may be given to members of designated groups in pursuing the numerical goals set out in employment equity plans, these goals may not be treated as rigid quotas. Section 15(4) amplifies this by providing that affirmative action measures may not translate into absolute barriers to the employment or promotion of members from non-designated groups. The implication of these requirements is that the employment equity plan must be treated as a target rather than a set of rigid requirements.

In taking the decision to refuse Barnard's promotion, the National Commissioner was also implementing the 2004 National Instruction.⁶⁴ The instruction made the employment equity plan binding on all members of the SAPS by requiring interviewing panels to pay regard to the employment equity plan, among other considerations, in making decisions. All promotions to salary levels 8 and above had to be forwarded to the National Commissioner for his final determination. The instruction also specified that the fact that a candidate is judged to be the best does not require promotion. Furthermore, the National Commissioner could leave positions vacant if deemed appropriate.⁶⁵

After failing to resolve the matter in the CCMA, Barnard instituted proceedings in the Labour Court. She was represented by Solidarity, a conservative trade union which actively opposes affirmative action and has brought numerous cases in an attempt to advance its cause.⁶⁶ That fact could not have been lost on the courts.

In her statement of claim, Barnard alleged that she had been unfairly discriminated against on the basis of race under s 6(1) of the EEA.⁶⁷ Section 6(1)

⁶¹ EEA, s 1 defines designated employers as including 'an organ of state as defined in s 239 of the Constitution, but excluding the National Defence Force, the National Intelligence Agency and the South African Secret Service'.

⁶² EEA, s 1.

⁶³ *Barnard* (note 2 above) at para 66.

⁶⁴ *Ibid* at para 46.

⁶⁵ *Ibid*.

⁶⁶ See note 15 above.

⁶⁷ *Barnard* (note 2 above) at para 18.

largely mirrors the prohibition on unfair discrimination under s 9 of the Constitution.⁶⁸ Section 11 of the EEA also replicates s 9(5) of the Constitution by imposing a presumption of unfairness where discrimination occurs on listed grounds. However, s 6(2)(a) expressly provides for affirmative action, stating that ‘it is not unfair discrimination to ... take affirmative action measures consistent with the purposes of the Act’. This provision had previously been interpreted as an interpretive guide in applying the test for unfair discrimination, rather than a standalone test for valid affirmative action.⁶⁹

In bringing her case to court, Barnard did not challenge the SAPS employment equity plan or the National Instruction. Her complaint was solely directed at the National Commissioner’s implementation of these measures in deciding not to appoint her to the vacant post.

A further issue was the paucity of evidence either for or against the National Commissioner’s decision. The National Commissioner did not testify in the Labour Court, nor did he depose to an affidavit explaining his reasoning.⁷⁰ It was left to a lower ranking official to testify in support of the National Commissioner’s reasons, although he did not have personal knowledge of the decision.⁷¹ Barnard also made little attempt to counter the National Commissioner’s reasons. In particular, she provided no concrete evidence to show that the National Commissioner’s decision compromised service delivery.⁷²

As a result, the courts were presented with the difficult task of assessing the validity of the implementation of affirmative action measures, rather than the measures themselves, with little evidence on which to form an assessment. This combination of the narrowness of the challenge and limited information were to be decisive in the decisions that followed.

B Lower Courts’ Decisions

Over a period of seven years, Barnard’s case made its way from the Labour Court to the Labour Appeal Court and then to the Supreme Court of Appeal.⁷³ The courts reached different conclusions on the merits of her claim but their decisions shared a common failure to apply the *Van Heerden* test. They differed on

⁶⁸ Section 6 of the EEA provides, in relevant part:

- ‘(1) No person may unfairly discriminate, directly or indirectly, against an employee, in any employment policy or practice, on one or more grounds, including race, gender, sex, pregnancy, marital status, family responsibility, ethnic or social origin, colour, sexual orientation, age, disability, religion, HIV status, conscience, belief, political opinion, culture, language, birth or on any other arbitrary ground.
- (2) It is not unfair discrimination to
 - (a) take affirmative action measures consistent with the purpose of this Act; or
 - (b) distinguish, exclude or prefer any person on the basis of an inherent requirement of a job.’

⁶⁹ See note 6 above.

⁷⁰ *Barnard* (note 2 above) at para 104.

⁷¹ SAPS written submissions at para 74.

⁷² *Barnard* (note 2 above) at para 122.

⁷³ For commentary on these lower court decisions, see M Mushariwa ‘Who Are the True Beneficiaries of Affirmative Action? *Solidarity obo Barnard v SAPS*’ (2011) *Obiter* 439; Pretorius ‘Standard of Review’ (note 9 above); M McGregor ‘Affirmative Action on Trial – Determining the Legitimacy and Fair Application of Remedial Measures’ (2013) *Tydskrif vir die Suid-Afrikaanse Reg* 650; Pretorius ‘Unresolved Search’ (note 9 above).

what to make of the limited reasoning and evidence in support of the National Commissioner's decision.

In the Labour Court, Pretorius AJ applied the *Harksen* test in considering whether the National Commissioner's implementation of the employment equity plan breached s 6(1) of the EEA.⁷⁴ He found that the National Commissioner had discriminated against Barnard on the basis of race and that the discrimination had to be presumed to be unfair, in accordance with section 11 of the Employment Equity Act.⁷⁵ Given the absence of reasoning or evidence, Pretorius AJ concluded that the presumption had not been rebutted.⁷⁶

The Labour Appeal Court also applied the *Harksen* test, but concluded that the discrimination was fair.⁷⁷ In doing so, it reduced the unfairness analysis to a rationality enquiry.⁷⁸ In applying this standard of review, the LAC also appeared to place the burden of proof on Barnard. It concluded that the National Commissioner's decision was rationally linked to the employment equity plan.

The Supreme Court of Appeal reversed the LAC's finding.⁷⁹ In doing so, it also relied on the *Harksen* test and invoked the presumption of unfairness.⁸⁰ Having found that there was discrimination on the basis of race, the SCA held that the paucity of reasoning and evidence to support the National Commissioner's decision meant that the decision had to be presumed to be unfair.⁸¹

V BARNARD: THE CONSTITUTIONAL COURT'S DECISION

The confusion in the lower courts over the proper approach to affirmative action required a firm response from the Constitutional Court.⁸² It also gave the Court its first opportunity in a decade to reaffirm the *Van Heerden* test, to explain its application in the employment context, and to refine this test. In particular, clarity was needed on the precise standard of review required and intensity with which it ought to be applied.

The majority judgment and three concurrences were unanimous in the result: the National Commissioner's implementation of the employment equity plan was valid. However, the result was always going to be the least interesting and important part of the judgment. It was the manner in which the Court reached the result that was significant, as it was hoped this would provide guidance for future affirmative action cases. In this respect, the members of the Court were united by the view that it was impermissible to apply the *Harksen* test for unfair discrimination to the National Commissioner's decision. Beyond this point of

⁷⁴ *Solidarity obo Barnard and Another v South African Police Services* [2010] ZALC 10, 2010 (10) BCLR 1094 (LC).

⁷⁵ *Ibid* at para 26.

⁷⁶ *Ibid* at paras 36, 43.6.

⁷⁷ *South African Police Services v Solidarity obo Barnard* [2012] ZALAC 31, 2013 (3) BCLR 320 (LAC) at para 21.

⁷⁸ *Ibid* at para 44.

⁷⁹ *Solidarity obo Barnard v South African Police Service* [2013] ZASCA 177, 2014 (2) SA 1 (SCA).

⁸⁰ *Ibid* at paras 50 and 76.

⁸¹ *Ibid* at para 76.

⁸² *Barnard* (note 2 above).

agreement, the different judgments approached the issues in very different ways, applying different standards of review.

A The Moseneke Majority and Jafta Concurrence

Moseneke ACJ was the architect of the Court's *Van Heerden* test. It was expected that he would take the opportunity to apply this test under the EEA and, in doing so, that he would clarify and expand on the appropriate intensity of review. However, he largely avoided these issues in his majority judgment.

This avoidance involved three steps. First, Moseneke ACJ held that the SCA erred by applying the *Harksen* test for unfair discrimination at the outset, without first assessing whether the National Commissioner's decision amounted to the valid implementation of an affirmative action measure.⁸³ In doing so, the SCA had misapplied the law.

Second, Moseneke ACJ then declined to set out and apply the correct test for assessing the validity of the implementation of affirmative action measures. This was because he characterised Barnard's case as having abandoned the unfair discrimination challenge.⁸⁴ Instead, he held that her case had morphed into a quasi-administrative law review of the rationality and reasonableness of the National Commissioner's decision.⁸⁵ He suggested that this challenge must fail as it was not properly pleaded.⁸⁶

Third, he indicated that even if the Court were to consider the merits of the new challenge it would fail as the National Commissioner's decision had not been shown to be irrational and unreasonable.⁸⁷ This conclusion followed principally because Barnard bore the burden of proof.

Albertyn describes the majority judgment as an 'oddly formalistic side-step' that avoided the main issues.⁸⁸ The majority was correct that the basis of Barnard's challenge in the Constitutional Court was obscure. However, as Cameron et al made clear, the essence of the complaint had always been one of unfair discrimination.⁸⁹ This necessarily required the Court to engage with the validity of the implementation of the employment equity plan. Indeed, Moseneke ACJ acknowledged the essence of the case in the opening paragraph of his judgment, where he noted '[t]he core issue in [the SCA] and this Court remains unchanged. Did the National Commissioner's decision unfairly discriminate against the respondent?'⁹⁰

Seen in this light, the overwhelming impression created by the majority judgment is of a Court searching for an easy way out of a difficult task. In doing so, the majority did not merely miss the opportunity to develop and clarify the *Van Heerden* test, it actively declined this opportunity where it could and should have seized it. Of course, judicial minimalism has its place in constitutional

⁸³ Ibid at paras 51–53.

⁸⁴ Ibid at paras 58–59.

⁸⁵ Ibid.

⁸⁶ Ibid at para 60.

⁸⁷ Ibid at para 61–70.

⁸⁸ Albertyn (note 8 above) 716.

⁸⁹ *Barnard* (note 2 above) at para 82.

⁹⁰ Ibid at para 1.

jurisprudence, but not in this case.⁹¹ More was required of the Court given the confusion about the appropriate intensity of review in affirmative action cases.

While the majority judgment lays down very little precedent, it does establish one principle and suggests another. First, Moseneke ACJ confirmed that the *Van Heerden* test does apply under the EEA, but only in respect of assessing the merits of employment equity plans and other affirmative action ‘measures’, not their implementation.⁹²

Second, he suggested that the implementation of affirmative action measures is to be assessed by a different test.⁹³ He hinted that this test requires legality and rationality, at minimum, but he refused to be drawn on the precise standard of review or the intensity with which it is to be applied. He concluded ‘these are the minimum requirements, it is not necessary to define the standard finally’.⁹⁴ These comments on assessing the implementation of affirmative action were strictly obiter, but they may be developed in future decisions.

In his separate concurrence, Jafta J also strongly suggested that rationality is the preferred standard for assessing the implementation of affirmative action measures.⁹⁵ However, he argued that the question of the appropriate standard of review had not been raised by any of the parties.⁹⁶ This was not strictly correct, as the heads of argument prepared by the SAPS argued that the *Van Heerden* test should be applied in assessing affirmative action measures *and* their implementation.⁹⁷ This again suggests a retreat from deciding matters that were squarely before the Court.

B The Cameron, Froneman and Majiedt Concurrence

The judgment of Cameron et al criticised the majority for avoiding the key issues. In their view, the issue of unfair discrimination had been squarely raised and it fell to the Court to decide on the appropriate standard of review:

This is the first case before this Court that deals with the standard to be applied in assessing the lawfulness of the individual implementation of constitutionally compliant restitutionary measures. It is important to give guidance on this difficult issue.⁹⁸

⁹¹ On the appropriate place of minimalism in judicial reasoning, see CR Sunstein *One Case at a Time: Judicial Minimalism on the Supreme Court* (1999) ch 1. For further discussion of maximalism and minimalism in the South African context (and the related, but not identical, distinction between substantive and formalist reasoning), see, eg, A Cockrell ‘Rainbow Jurisprudence’ (1996) 12 *South African Journal on Human Rights* 1; I Currie ‘Judicious Avoidance’ (1999) 15 *South African Journal on Human Rights* 138; C Roederer ‘Judicious Engagement: Theory, Attitude and Community’ (1999) 15 *South African Law Journal* 486; S Woolman ‘The Amazing, Vanishing Bill of Rights’ (2007) 124 *South African Law Journal* 762; K Young ‘The Avoidance of Substance in Constitutional Rights’ (2015) 5 *Constitutional Court Review* 233.

⁹² *Barnard* (note 2 above) at para 38.

⁹³ *Ibid* at para 38–39.

⁹⁴ *Ibid* at para 39.

⁹⁵ *Ibid* at paras 227–229.

⁹⁶ *Ibid* at para 219.

⁹⁷ SAPS written submissions at paras 63–68, see para 67 in particular.

⁹⁸ *Barnard* (note 2 above) at para 75.

Cameron et al proceeded to argue that the appropriate standard of review for assessing the implementation of affirmative action measures under the EEA is ‘fairness’.⁹⁹ In doing so, they appeared to agree with the majority that the *Van Heerden* test should not be applied to implementation. Precisely why they took this stance is not fully explained in the judgment. In a single paragraph, they suggest that unfair discrimination under the Constitution involves different considerations and problems to unfair discrimination under the EEA, necessitating a different approach to affirmative action.¹⁰⁰ However, that justification overlooks the fact that the *Harksen* test for unfair discrimination has been consistently applied under s 6(1) of the EEA, even though it was first developed under s 9(3) of the Constitution. There is no apparent reason why the *Van Heerden* test under s 9(2) cannot follow suit and be applied under the EEA, with some necessary tailoring to the context.

Cameron et al also found themselves with the difficult task of explaining how a test of ‘fairness’ in the context of affirmative action measures is different to a test for the fairness of discrimination. The same word can, of course, be used to describe different concepts.¹⁰¹ However, few will be so astute as to notice the difference, creating great potential for confusion.

The precise nature of the ‘fairness’ standard of review was not fully explained in the judgment. However, two features of this test were made clear. First, Cameron et al firmly indicated that ‘fairness’ involves more than rationality. In their view, rationality is too deferential to allow the courts to interrogate the balance between the benefits and harms of affirmative action measures. See, for instance, the following statement:

The important constitutional values that can be in tension when a decision-maker implements remedial measures require a court to examine this implementation with a more exacting level of scrutiny.¹⁰²

And further:

[A] rationality standard does not allow a court to interrogate properly a decision-maker’s balancing of the multiple designated groups, or of their interests against those adversely affected by the restitutionary measures.¹⁰³

We must therefore formulate a standard specific to the Act, one that is rigorous enough to ensure that the implementation of a remedial measure is “consistent with the purpose of [the] Act” – namely, to avoid over-rigid implementation, to balance the interests of the various designated groups, and to respect the dignity of rejected applicants.¹⁰⁴

Second, Cameron et al indicated that the more determinate rules on affirmative action in the EEA are bundled up in the fairness standard.¹⁰⁵ On this approach,

⁹⁹ Ibid at para 101.

¹⁰⁰ Ibid at para 97.

¹⁰¹ See, eg, *S v Dlamini* [1999] ZACC 8, 1999 (4) SA 623 (CC), 1999 (7) BCLR 771 (CC) at para 47 on the different meanings of ‘interests of justice’.

¹⁰² *Barnard* (note 2 above) at para 95.

¹⁰³ Ibid at para 96.

¹⁰⁴ Ibid at para 97.

¹⁰⁵ Ibid at para 100.

the prohibitions on quotas and job reservations under ss 15(3) and (4) of the EEA appear to fall within this all-things-considered fairness analysis.

In applying the fairness standard to the facts, Cameron et al gave some hint that this standard could be applied with variable intensity in different cases. The central question they confronted was whether the paucity of evidence and reasoning in support of the National Commissioner's decision was fatal. Cameron et al used a number of techniques to find that the balance ultimately weighed in the National Commissioner's favour.

First, the burden of proof was placed squarely on Barnard. It was clear from the way that Cameron et al considered the matter that the applicant bore the risk of non-persuasion and that the benefit of any doubt would be given to the National Commissioner.¹⁰⁶

Second, Barnard also carried the evidential and argumentative burden. Cameron et al emphasised that, in the circumstances, Barnard had to make out a case that her promotion was in the interests of service delivery, that gender representation would be advanced by her promotion, and that these interests outweighed the interests of promoting racial representation at salary level 9. The Court was not going to come to Barnard's assistance in making these arguments and providing evidence in support of them. However, Cameron et al did appear to suggest that if Barnard had based her case on gender rather than race, then the outcome could have been different.¹⁰⁷ Nonetheless, they held that '[i]n the absence of proper challenge and argument, the Court cannot undercut the decision-maker's stated reasons on this point'.¹⁰⁸

Finally, Cameron et al came to the assistance of the National Commissioner by going beyond the limited reasons he provided for the decision. While the National Commissioner failed to provide any real justification, Cameron et al held that 'this is not fatal, because there are sufficient external facts to determine that the National Commissioner's decision was fair'.¹⁰⁹ In this way, fairness can be determined on the objective facts, rather than the justification offered by the decision-maker. This approach is summed up in the final paragraphs of the judgment:

We conclude that the facts show that the National Commissioner's decision passes the fairness standard. While we find this is a close call, what has proved determinative to us is the pronounced over-representation of white women at the salary-level to which Ms Barnard was applying. This was not just one or two, but many. There was thus greater justification for prioritising racial representivity over other considerations.¹¹⁰

This approach is not only deferential to the National Commissioner, in the sense of affording greater weight to his reasons than they would otherwise carry. It involves the Court actively seeking out evidence and reasons in favour of the decision. This approach may well be justified in the circumstances of the case.

¹⁰⁶ Ibid at para 122.

¹⁰⁷ Ibid at para 115.

¹⁰⁸ Ibid at para 122.

¹⁰⁹ Ibid at para 104.

¹¹⁰ Ibid at para 123.

However, no explicit justification was given for applying the standard of review in this way.

The great point of uncertainty in this judgment is when and how the intensity of review may vary and for what reasons. Cameron et al were largely silent on this issue. At various points they signal that the way they have applied the evidential and argumentative burden may be different in other cases. In particular, they implied that if Barnard had made gender representation the focus of her case, then the Court would have applied a more intense level of scrutiny. They also strongly emphasised the importance of state accountability and reason-giving, suggesting that in appropriate circumstances the state's silence in defending specific instances of affirmative action may count against it.¹¹¹ However, no concrete guidance was offered for future decisions.

C The Van der Westhuizen Concurrence

In his separate concurrence, Van der Westhuizen J was the only member of the Court to apply the *Van Heerden* test in assessing the validity of the National Commissioner's decision.¹¹² In doing so, he also rejected the view that rationality suffices.¹¹³

While Van der Westhuizen J applied the *Van Heerden* test, he did not interpret it as incorporating a proportionality analysis. Instead, he suggested that this proportionality analysis is located outside the *Van Heerden* test.¹¹⁴ What is never explained in the judgment is why the balancing of interests cannot be done within the third leg of the *Van Heerden* enquiry, as the *Van Heerden* Court appeared to require.

In applying this analysis, Van der Westhuizen J proceeded to determine whether the National Commissioner's decision impacted on Barnard's dignity and undermined service delivery in a manner that was disproportionate.

Van der Westhuizen J placed the burden of proof squarely on Barnard. He also followed Cameron et al in considering evidence and arguments in favour of the National Commissioner's decision which had not been advanced by the National Commissioner himself. This approach is captured in his discussion of the difficulties of assessing the impact of the decision on service delivery:

Without proper evidence or specialist institutional knowledge, it may be difficult for a court to draw conclusions about the precise impact a policy, an appointment, or even a vacancy will have on service delivery. This is the reason for the National Commissioner's wide discretionary powers, particularly in the context of affirmative measures, to appoint a candidate or keep a post vacant. In this case there is not enough evidence for this Court to impugn the decision on the issue of service delivery. It cannot be said that it was disproportionate for the National Commissioner to rank representivity higher than the possible impact on service delivery in this case.¹¹⁵

¹¹¹ Ibid at paras 103–107.

¹¹² Ibid at paras 133, 142–143.

¹¹³ Ibid at para 141.

¹¹⁴ Ibid at para 165–167.

¹¹⁵ Ibid at para 189.

Like Cameron et al, there is little in the way of open justification for applying the proportionality analysis in this way. The justification is hinted at rather than addressed directly. That justification lies necessarily in the nature of the interests at stake, the court's institutional competence on these issues, and questions of the Court's legitimacy in this area. However these reasons were largely obscured by talk of dignity, equality and other values.

D Missed Opportunities

In sum, the *Barnard* Court did not provide definitive answers to the three questions raised by *Van Heerden*. This is likely to cause difficulties for courts confronted with affirmative action cases in future.

First, there is likely to be further confusion about the application of the *Van Heerden* test under s 6 of the EEA to the implementation of affirmative action measures. The one point on which the *Barnard* majority established precedent is that the *Van Heerden* test applies when considering the validity of affirmative action measures, such as employment equity plans. However, it remains uncertain whether the *Van Heerden* test applies to the implementation of these affirmative action measures. The majority judgment and the Cameron et al and Jafta concurrences all suggest that the implementation of affirmative action will be assessed by some other test, although no clarity is provided on what this should be.

Second, the divergent judgments offer no concrete guidance on the appropriate standard of review when assessing affirmative action measures. The majority judgment suggests an inclination to rationality, although this was not finally determined. Only Jafta J was firmly in favour of making rationality the exclusive standard of review. Cameron et al and Van der Westhuizen J favoured a proportionality analysis, involving the balancing of interests, although both located this analysis outside of the *Van Heerden* framework.

Third, there was no clear guidance on when and how these standards of review may be applied with varying intensity. This is significant as the judges' chosen intensity of review in applying their different standards proved to be the decisive factor in the case. Albertyn notes that it is surprising that the judges could reach the same result despite using such different standards of review.¹¹⁶ The result is not at all surprising, however, when we recognise that the majority and concurring judgments applied different standards of review with the same, low intensity. In particular, all of the judgments placed the burden of proof squarely on the applicant. Once the Court accepted that *Barnard* had the burden to prove that the implementation of the affirmative action measures was invalid, the result was a foregone conclusion, no matter what standard of review was applied. *Barnard*'s failure to put up real evidence or argument to challenge the National Commissioner's reasons, as scant as they were, meant that her claim could not succeed. The choice to place the burden of proof on *Barnard* was justified, for reasons I will expand on below. However, the members of the Court did not attempt to justify this choice or even acknowledge that there was a choice to be

¹¹⁶ Albertyn (note 8 above) 722.

made. That choice was central to the outcome of the case and required greater candour and explanation.

After a wait of ten years, *Barnard* did not deliver the expected guidance on these issues. The Constitutional Court did not, however, have to wait long for another affirmative action case. In its 2016 judgment in *Solidarity v Department of Correctional Services* the Court was again confronted with an affirmative action case; however the narrow scope of the arguments did not allow the Court to address the questions left open in *Barnard*.¹¹⁷

Solidarity concerned a direct challenge to the Department of Correctional Services' employment equity plan in the Western Cape. This was in contrast with *Barnard*, where the plan was not challenged. Ten prison warders, supported by *Solidarity*, challenged the Department's refusal to appoint them to various positions. Nine of the ten identified as Coloured and several were women. They were denied appointment on the basis that Coloured people and women were overrepresented in the relevant posts, according to the targets set in the employment equity plan. It was common cause that these targets were based on national demographics and did not take account of the specific demographics of the Western Cape. The Court found that s 42(a) of the EEA, as it was then worded, required due consideration of regional demographics.¹¹⁸ As a result, the Court held that the plan was in breach of s 42(a) and the refusal to appoint the warders on the basis of this flawed plan was therefore invalid.¹¹⁹

The Court was not required to reach any firm findings on the intensity of review in affirmative action cases. Nevertheless, the Court did lay down three important principles. First, the Court confirmed what it called the '*Barnard* principle': that the *Van Heerden* test applies when assessing the validity of affirmative action measures, including employment equity plans.¹²⁰ Second, the Court held that the *Van Heerden* test applies even when considering claims brought by historically disadvantaged individuals who have been denied benefits under an affirmative action scheme.¹²¹ This confirms that the *Van Heerden* test applies irrespective of the identity of the party challenging the affirmative action measure. Thirdly, the majority also rejected the argument that the employment equity plan was a disguised quota, which is prohibited under s 15(3) of the EEA.

Given the narrow scope of this decision, the Court did not engage further with the nature of the *Van Heerden* test and the appropriate intensity of review in applying this test. The Court also shed no further light on the appropriate standard of review in assessing the implementation of affirmative action measures, as it was primarily concerned with the employment equity plan itself

¹¹⁷ *Solidarity and Others v Department of Correctional Services and Others* [2016] ZACC 18, 2016 (5) SA 594 (CC), 2016 (10) BCLR 1349 (CC) ('*Solidarity*').

¹¹⁸ *Ibid* at para 74. At the relevant time, s 42(a) of the EEA provided that in assessing compliance with the EEA, the relevant bodies 'must' take into account the 'demographic profile of the national and regional economically active population'. Section 42(a) was subsequently amended to replace 'must' with 'may'.

¹¹⁹ *Ibid* at para 82.

¹²⁰ *Ibid* at para 25, summarising and endorsing the Labour Appeal Court's interpretation of *Barnard* in *Solidarity and Others v Department of Correctional Services and Others* [2015] ZALAC 6, 2015 (4) SA 277 (LAC) at para 51.

¹²¹ *Ibid* at paras 37–49.

and not its implementation. As a result, the questions left open in *Barnard* remain unanswered.

VI DEVELOPING THE *VAN HEERDEN* TEST IN FUTURE DECISIONS

I now turn to give arguments showing how the Constitutional Court can refine the *Van Heerden* test in future decisions. Three key developments are required, which I elaborate upon below.

A The Application of *Van Heerden* under the EEA

The first step, I contend, is that the Constitutional Court should apply the *Van Heerden* test both to affirmative action measures *and* their implementation.

It must be acknowledged that there is a conceptual and analytical distinction worth making between affirmative action measures and their application in practice. An employment equity plan may be valid on its face but may be applied in ways that are invalid. An abstract assessment of an employment equity measure may also involve different factual considerations than a detailed assessment of how the affirmative action measure is being applied in practice.

However, the fact that there is a difference between an affirmative measure and its implementation does not mean that different tests are required to assess their validity. There is no principled or practical reason for creating entirely separate tests.

First, as I noted earlier, it is difficult to see why two separate tests are needed for the validity of affirmative action measures and their implementation when one test suffices for unfair discrimination cases. There is no bifurcation of tests in assessing unfair discrimination challenges under section 6 of the EEA. The *Harksen* test is used to assess workplace policies and their day-to-day implementation. Precisely why the Court considered it necessary to have two separate tests for the validity of affirmative action measures was never adequately explained in *Barnard*.

Second, as Van der Westhuizen J pointed out, the *Van Heerden* test was clearly designed to consider both affirmative action measures and their implementation.¹²² This is evident in the third requirement, which focuses on the benefits of the affirmative action measure and its negative impact on those who are excluded. It would be impossible to consider the beneficial and harmful impact of affirmative action measures without any regard for how they are or will be implemented in practice. Assessing the validity of measures and their implementation in isolation would result in an impossible abstraction for a test which is meant to be contextualised in reality.¹²³

Third, and most importantly, separate tests may dilute the protections afforded to historically disadvantaged groups. This is because employment equity plans and their implementation both have the potential to harm disadvantaged groups to the same extent, and so require the same standard of review. For example, consider an employment equity plan that sets ambitious targets for the promotion

¹²² *Barnard* (note 2 above) at para 143.

¹²³ See *Van Heerden* (note 3 above) at paras 44, 142.

of disabled people but is implemented in ways that ignore the interests of this group. Compare this with an employment equity plan that, in its design, gives insufficient weight to the interests of disabled people. Both scenarios would likely impact and exclude disabled people in the same way. As a result, it would make no sense to subject the exclusionary plan to a more intense standard of review than the improper implementation of the otherwise inclusive plan. Doing so would risk allowing abuses and undue burdens to slip through without sufficient checks.

The application of the *Van Heerden* test under the EEA is not without challenges, which will require further thought as the case law develops. In particular, a clearer account is needed of how the *Van Heerden* test relates to the more determinate provisions in the Act. For example, the Act contains rule-like norms on regional representivity,¹²⁴ a prohibition on quotas and absolute job reservations,¹²⁵ and requirements that candidates be suitably qualified.¹²⁶ The judgments of Van der Westhuizen J and Cameron et al appear to suggest that these rule-like norms should be bundled into the broader balancing enquiry. This does not appear to be appropriate. It is the nature of rules that they are meant to be determinative of specific factual situations, providing more precise guidance to decision-makers.¹²⁷ Placing these rule-like norms into an all-things-considered balancing enquiry may dilute their force. As a result, it would be better to consider these rules as binding constraints that are independent of the *Van Heerden* test. Compliance with these rules is a necessary condition for the validity of an affirmative action measure, but it is not sufficient to guarantee such validity. An affirmative action measure that complies with these rules could still be invalidated under the *Van Heerden* test if it does not strike an appropriate balance between the competing interests at stake. For example, an affirmative action measure may not be a prohibited quota, but it may still fail the *Van Heerden* test.

B The Standard of Review

The next development that is needed is confirmation that the *Van Heerden* test involves a proportionality analysis, rather than a mere rationality test. As Albertyn notes, Van der Westhuizen J was on the right track in this respect.¹²⁸ However, he curiously decided to locate this proportionality analysis outside of the *Van Heerden* test. That is unnecessary as the third requirement of this test is sufficiently flexible to accommodate this analysis.

The judgment of Cameron et al in *Barnard* provides the strongest justification for why a rationality analysis does not suffice.¹²⁹ If the test for valid affirmative action was solely concerned with the connection between ends and means then this would overlook the potential negative impact that affirmative action measures can have. All that would be required is proof that the measure has some

¹²⁴ EEA, s 42.

¹²⁵ EEA, s 15(3) and (4).

¹²⁶ EEA, s 15(1).

¹²⁷ On the distinction between rules, standards, principles and other norms, see CR Sunstein 'Problems with Rules' (1995) 3 *California Law Review* 953. See also Raz (note 1 above) ch 1.

¹²⁸ Albertyn (note 8 above) 730.

¹²⁹ *Barnard* (note 2 above) at paras 94–96.

rational connection to the advancement of a disadvantaged group, ignoring any impact that this measure may have on other groups.

Overlooking the possible negative consequences of affirmative action would be especially problematic where members of historically disadvantaged groups are excluded or actively harmed by these measures. The Constitutional Court's most recent judgment in *Solidarity* now confirms that the *Van Heerden* test will apply even where disadvantaged groups are harmed or excluded. If the impact of the affirmative measure on these groups was not taken into account, this could have serious consequences. To take a stylised example, consider an affirmative action measure that creates bursaries for women to attend university by reallocating all of the budget that was previously assigned for bursaries that were to be awarded to disabled students. A pure test of rationality would not allow any consideration of the impact that this measure may have on the disabled students. It would only be concerned with assessing whether the measure is rationally capable of advancing the legitimate purpose of benefitting women students. There would be no scope for a court to balance the benefits of this measure against the harms to disabled bursary recipients, nor would there be any scope for a court to assess the availability of less restrictive alternatives, such as finding alternative sources of funding to allow bursaries for both groups. Those questions would require a proportionality analysis.

Proportionality assesses whether the benefits outweigh the costs in each case. Proportionality analysis is most commonly associated with the s 36 limitations analysis under the Constitution, but it is not the sole preserve of s 36. Proportionality is applied in areas as diverse as the test for arbitrary deprivations of property,¹³⁰ unfair discrimination,¹³¹ and public participation in law-making,¹³² among others.

Proportionality analysis is also not exclusive to the legal domain. As Rivers argues, proportionality is a tool of practical reasoning that is applied whenever we deliberate about the correct course of action in the face of competing interests and scarce resources:

The doctrine of proportionality is not simply a legal device to assist judges in regulating legislative and executive incursions on rights. It better understood as a rational device for the optimisation of interests.¹³³

The logical structure of proportionality remains the same wherever it is applied, in law or in other decision-making contexts. What differs is the intensity and rigour with which it is applied. As a result, proportionality must be coupled with

¹³⁰ *First National Bank of SA Ltd t/a Wesbank v Commissioner, South African Revenue Service and Another; First National Bank of SA Ltd t/a Wesbank v Minister of Finance* [2002] ZACC 5, 2002 (4) SA 768 (CC), 2002 (7) BCLR 702 (CC) at para 86 (a deprivation of property is arbitrary if it is disproportionate to its purpose, albeit that this is less exacting than a s 36 analysis).

¹³¹ See authors cited at notes 36–37 above.

¹³² *Doctors for Life International v Speaker of the National Assembly and Others* [2006] ZACC 11, 2006 (6) SA 416 (CC), 2006 (12) BCLR 1399 (CC) at para 128 (the test for whether the legislature has acted reasonably in facilitating public participation depends on a balancing of multiple considerations, bearing strong resemblance to a proportionality analysis).

¹³³ Rivers (note 19 above) 207.

a set of principles and other norms to govern the variable intensity of review.¹³⁴ I will say more about these norms in a moment.

Before exploring the ways in which the proportionality analysis can be varied, it is important to understand the logical structure of the proportionality analysis. Where the Constitutional Court has applied a proportionality analysis in other areas of its jurisprudence, it has avoided a ‘check-list’ approach to proportionality, as is favoured in other jurisdictions.¹³⁵ A check-list approach involves treating the proportionality analysis as a list of requirements that must be ticked off in sequence.¹³⁶ Instead the Constitutional Court opts for what it terms a ‘global’ judgment on proportionality, meaning that it reaches an all-things-considered judgment on where the balance of interests lies.¹³⁷

While the Court does not approach proportionality in a structured way, the common features of proportionality analysis are reflected in its reasoning.¹³⁸ These are the questions of whether the law or action has a legitimate purpose (*legitimacy*), whether it is capable of achieving that purpose (*suitability*), whether there are less restrictive means available to achieve that purpose (*necessity*), and, on balance, whether the attainment of that purpose outweighs the costs (*balancing*).¹³⁹

The three requirements of the *Van Heerden* test incorporate these elements of the proportionality analysis. The first two requirements reflect the dimensions of legitimacy and suitability. The last requirement involves balancing the competing interests. This balancing would necessarily include some assessment of necessity: whether there are other, less restrictive means to achieve the purpose of protecting or advancing the disadvantaged group. If an affirmative action measure could be achieved in other ways that are less exclusionary or harmful to others, then this is a factor to weigh in the balance.¹⁴⁰

On first glance, the majority judgment in *Van Heerden* appeared to rule out any consideration of necessity. This is reflected in the following passage:

The provisions of s 9(2) do not prescribe ... a necessity test because remedial measures must be constructed to protect or advance a disadvantaged group. They are not predicated on a necessity or purpose to prejudice or penalise others, and so require supporters of the

¹³⁴ Ibid 182.

¹³⁵ On structured proportionality analysis, see generally A Barak *Proportionality: Constitutional Rights and Their Limits* (2012) chs 9–12.

¹³⁶ See, eg, the ‘*Oakes* test’ for justifiable limitations under s 1 of the Canadian Charter, developed in *R v Oakes* [1986] 1 SCR 103.

¹³⁷ *Manamela* (note 25 above) at para 32. See N Petersen ‘Proportionality and the Incommensurability Challenge in the Jurisprudence of the South African Constitutional Court’ (2014) 30 *South African Journal on Human Rights* 405, 407.

¹³⁸ Petersen (note 137 above). While proportionality analyses share common features, courts in different jurisdictions adopt different approaches to this analysis. See further D Grimm ‘Proportionality in Canadian and German Constitutional Jurisprudence’ (2007) 57 *University of Toronto Law Journal* 383.

¹³⁹ For deeper analysis of these components of the proportionality analysis, see Barak (note 135 above) chs 6, 9–12.

¹⁴⁰ For further examination of this necessity component of the proportionality analysis, and the different forms that necessity analysis can take, see D Bilchitz ‘Necessity and Proportionality: Towards a Balanced Approach?’ in L Lazarus, C McCrudden & N Bowles (eds) *Reasoning Rights* (2014); Petersen (note 137 above).

measure to establish that there is no less onerous way in which the remedial objective may be achieved.¹⁴¹

The point Moseneke J made in this passage is that the validity of affirmative action does not require proof that there is no other way to protect or advance the disadvantaged group. In doing so, Moseneke J does not claim that consideration of less restrictive alternatives is irrelevant to the enquiry. He simply emphasises that the availability of less restrictive alternatives is not determinative of the validity of an affirmative action measure. In doing so, he is rejecting a test of ‘strict’ necessity.¹⁴² However, Moseneke J should not be interpreted as dispensing with any consideration of less restrictive alternatives. The availability of alternatives will not be conclusive, but it is a relevant consideration that must weigh in the balance.

While necessity forms a part of this test, courts will generally approach this enquiry with some degree of deference to the initial decision-maker. As Sachs J cautioned in *Van Heerden*:

Courts must be reluctant to interfere with such measures, and exercise due restraint when tempted to interpose themselves as arbiters as to whether the measure could have been proceeded with in a better or less onerous way.¹⁴³

As with all other elements of the *Van Heerden* test, the intensity with which a court evaluates the existence of less restrictive alternatives will depend on the context. I now turn to discuss this variable intensity of review in more detail.

C Taming the Variable Intensity of Review

By affirming that the *Van Heerden* test applies under the EEA and that the test involves a proportionality analysis, the Constitutional Court would go a long way in providing further clarity and guidance for lower courts. As *Barnard* shows, further guidance is also urgently needed on the questions of when and how this intensity of review may be varied in applying the *Van Heerden* test.

Variability in the intensity of review is necessary and unavoidable. The complexities of affirmative action, the different interests at play, and the different contexts in which affirmative action may be applied require some flexibility in the test. In particular, the intensity of review will need to be increased significantly where members of historically disadvantaged groups challenge affirmative action measures that exclude them or compound their disadvantage.

Two things are needed to bring greater guidance and accountability to this variable intensity of review. First, the Court must be conscious and open about the different ways in which the *Van Heerden* test can be applied with variable intensity. Second, it is necessary for the Court to articulate and apply a set of principles and other norms to justify why it has chosen a particular intensity of review in each case.

¹⁴¹ *Van Heerden* (note 3 above) at para 43.

¹⁴² See further Bilchitz (note 140 above).

¹⁴³ *Ibid* 152.

1 *The Points of Variability*

The three stages of the *Van Heerden* test each allow for variable intensity in their application. The first stage requirement – that historically disadvantaged groups should be in the ‘overwhelming majority’ of beneficiaries of an affirmative action measure – can be applied more or less strictly depending on the circumstances. This was evident in the disagreement between the majority judgment and the separate concurrences in *Van Heerden*. The evidence showed that 79 per cent of the beneficiaries of the affirmative action measure were black. Moseneke J, writing for the majority, concluded that this was an overwhelming majority. Mokgoro J and Ngobo J held that it was not. What was missing from these decisions was any justification for these different degrees of strictness in applying this requirement.

The second stage requirement – a reasonable likelihood that the affirmative action measure will benefit the disadvantaged group – is also capable of different degrees of scrutiny. In *Van Heerden*, Moseneke J stressed that this requirement does not necessarily require evidence showing that affirmative action actually produces benefits. As discussed above, this evidence will often be hard to come by, given that the long-term effects of affirmative action are often unknowable or hard to predict. Nevertheless, it is notionally possible that courts may vary this requirement depending on the case. If a court is confronted with an affirmative action measure that has a proven negative impact on vulnerable groups it would, in all likelihood, not be satisfied by vague and unsubstantiated predictions about the long-term benefits of the programme.

This second stage requirement is likely to be applied more strictly where the notional beneficiaries of an affirmative action programme are also its victims. Some affirmative action programmes may impose burdens and disadvantages on those they are intended to advance, in a way that entrenches patterns of disadvantage rather than removing them. The recent controversy over the UThukela ‘maidens’ bursary programme is a case in point.¹⁴⁴ A municipality provided bursaries to young women for their tertiary education, on the condition that they remain virgins and subject themselves to invasive virginity testing. This was arguably an affirmative action programme, aimed at providing predominantly black, rural women with educational opportunities. However, the conditions attached to these bursaries reinforced the stigma and double-standards surrounding women’s sexuality, as well as violating the privacy and bodily integrity of its beneficiaries. Measures like this are burdens disguised as benefits. Courts would and should subject the alleged benefits of such measures to more intense scrutiny.

The third requirement of the *Van Heerden* test is perhaps most open to variation. As was just explained, the proportionality analysis has numerous components. Each of these components can be applied more or less stringently in assessing where the balance of interests lies. In particular, a court’s willingness to interrogate the availability of less restrictive alternatives is likely to vary from case to case. In *Barnard* and *Van Heerden*, the Court was reluctant to interrogate alternatives. By contrast, if a court was confronted with a programme such as the UThukela

¹⁴⁴ See further M Smuts ‘State Sponsored Virginity: South Africa’s Maidens’ Bursary Scheme’ *Oxford Human Rights Hub Blog*, available at <http://ohrh.law.ox.ac.uk/state-sponsored-virginity-south-africas-maidens-bursary-scheme/>.

maidens' bursary then the availability of alternative means would hopefully weigh heavily in the analysis. These bursaries could still be offered to young women without the sexist and invasive conditions attached. The need to protect against harmful gender discrimination disguised as a benefit would require more intense scrutiny of these measures.

Another way in which this test can be applied more or less stringently is by adjusting the burden of proof. As indicated above, the burden of proof can be broken down into two elements: the burden of 'non-persuasion' and the evidential and argumentative burden.¹⁴⁵

The burden of non-persuasion entails that where the probabilities are evenly balanced or uncertain, the party bearing the burden will lose.¹⁴⁶ It is clear from the Court's decisions in *Van Heerden* and *Barnard* that the claimant in affirmative action cases will generally bear this burden. However, there is nothing preventing a court from placing this burden on the opposing party in appropriate cases.

The evidential and argumentative burden involves the responsibility to adduce evidence and argument on a point. As the Court recognised in *Mobunram v NDPP*,¹⁴⁷ the evidential and argumentative burden may be shifted, particularly where certain information is 'peculiarly within the knowledge' of one of the parties.

A court also has a choice to come to the assistance of one of the parties where the evidence and reasons provided are not sufficient to discharge the burden. This may involve the court taking judicial notice of facts not on the record or delving into the record to find evidence that supports or contradicts claims made by the decision-maker. It can also involve a court raising points of law *mero motu* either in support or against the affirmative action measure. These are all critical choices. Cameron et al and Van der Westhuizen J hinted that if Barnard had staked her claim on gender discrimination then the Court may have been more inclined to take a proactive role in finding evidence and arguments in her favour. In future cases, these choices may make all the difference to affirmative action measures, particularly where disadvantaged groups attack the under-inclusiveness or heavy-handedness of such measures.

There are many other subtle choices that may be made in adjusting the intensity of review in applying the *Van Heerden* test. These choices involve small nuances in the way that rules of procedure, evidence and doctrine are applied. Each of these choices has significance for the outcome of the case and ought to be guided by deeper principles. I now turn to explain these principles.

¹⁴⁵ Barak (note 137) 437; PJ Schwikkard 'The Standard and Burden of Proof and Evidential Duties in Civil Trials' in PJ Schwikkard & others (eds) *Principles of Evidence* (4th Edition, 2015) 616–617. These principles were articulated in *Pillay v Krishna* 1946 AD 946 at 952–953; *South Cape Corporation (Pty) Ltd v Engineering Management Services (Pty) Ltd* 1977 (3) SA 534 (A) at 548A–C.

¹⁴⁶ Chan (note 51 above) 47.

¹⁴⁷ *Mobunram and Another v National Director of Public Prosecutions and Another* [2007] ZACC 4, 2007 (4) SA 222 (CC), 2007 (6) BCLR 575 (CC) at para 75 (Court considered whether the forfeiture provisions under the Prevention of Organised Crime Act 121 of 1998 had been applied correctly).

2 Identifying the Principles

Variable intensity of review is nothing new to South African constitutional law. Commentators have often noted the unexplained variations in the intensity with which the Constitutional Court reviews state action in other areas of its jurisprudence.¹⁴⁸ As a result, there have been many calls for the Court to adopt a set of norms to govern this variable intensity of review; this could be termed a doctrine of deference. This doctrine aims to identify when it is appropriate for courts to defer to the state by adjusting the intensity of review.¹⁴⁹

Commentators have proposed a broad range of principles that should form part of this doctrine. In doing so, they have drawn on the more advanced UK and Canadian jurisprudence on deference.¹⁵⁰ The principles that emerge from these contributions can be grouped under three broad headings: the *interests at stake*, relative *democratic legitimacy*, and relative *institutional competence*.¹⁵¹

In affirmative action cases, the first set of principles, the *interests at stake*, will likely carry substantial weight in deciding how to vary the intensity of review. These principles operate at two levels. First, there is an appreciation of the importance and need for affirmative action measures in general, as affirmed by s 9(2) of the Constitution. This broad constitutional preference will generally tilt the scales in favour of affirmative action measures. Second, there is an appreciation of the specific interests at stake in each case. Where an affirmative action programme with proven benefits to a disfavoured group is under challenge, a court should be more circumspect in reviewing the justifications for this programme, requiring a lower intensity of review, all other things remaining equal. However, where there is an indication that the affirmative action programme has a severe impact, particularly on a historically disadvantaged group, then a court will be justified in making use of the flexibility in the *Van Heerden* test to scrutinise the affirmative action more closely. An intersectional approach is needed here that acknowledges the ways in which different forms of advantage and disadvantage connect and

¹⁴⁸ In the context of legality review, see authors cited at note 28 above. In the context of socio-economic rights, see K McLean *Constitutional Deference, Courts and Socio-Economic Rights in South Africa* (2009); A Pillay 'Economic and Social Rights Adjudication: Developing Principles of Judicial Restraint in South Africa and the United Kingdom' [2013] *Public Law* 599. For further analysis of the Constitutional Court's fluctuating intensity of review in other areas of its case law, see T Roux *The Politics of Principle: The First South African Constitutional Court, 1995-2005* (2013) chs 6–10.

¹⁴⁹ As discussed above, deference and intensity of review are generally inversely related. The greater the degree of deference shown by the court, the lower the intensity of review.

¹⁵⁰ See McLean (note 148 above) ch 1. On the UK's deference jurisprudence, see, eg, Rivers (note 19 above); Kavanagh 'Deference or Defiance?' (note 1 above); J King *Judging Social Rights* (2012). On the Canadian jurisprudence, see, eg, S Choudhry 'So What Is the Real Legacy of *Oakes*? Two Decades of Proportionality Analysis under the Canadian Charter's Section 1' (2006) 34 *Supreme Court Law Review* 501; R Dixon 'The Supreme Court of Canada, Charter Dialogue and Deference' (2009) 47 *Osgoode Hall Law Journal* 235.

¹⁵¹ McLean (note 148 above) chs 1, 2; A Price 'The Content and Justification of Rationality Review' in S Woolman & D Bilchitz (eds) *Is This Seat Taken? Conversations at the Bar, the Bench and the Academy about the South African Constitution* (2012) 50; Hoexter, *Administrative Law* (note 24 above) 151–155; Pillay (note 150 above) 600. This is by no means an exhaustive list, as other principles may also be relevant, such as the availability of non-judicial accountability mechanisms, polycentricity, comity and so on.

overlap.¹⁵² Black women, for example, experience the combined effects of racism and sexism, which create deeper and unique forms of disadvantage which are not shared by black men or white women. In many cases, these intersecting patterns of disadvantage will make stronger demands for redress. For instance, an affirmative action measure that harms or excludes black women would, in general, require more intense scrutiny than one that excludes white women.

Second, concern for *democratic legitimacy* requires courts to recognise that there are some matters on which unelected judges should be cautious about second-guessing the decisions of representative bodies. For example, an employment equity plan that has been developed in a bargaining council, involving representatives of trade unions and employers, will generally be reflective of the particular needs and interests of participants in the industry. A court should be cautious in second-guessing the delicate compromises that have been struck in these forums. However, in some cases democratic legitimacy may necessitate closer scrutiny, particularly where processes of deliberation have failed or have systemically ignored the interests of marginalised groups.¹⁵³ Furthermore, as the Court has stressed, democracy requires the development of a culture of justification, a culture that can be promoted by more searching scrutiny in appropriate cases.¹⁵⁴

Finally, considerations of *relative institutional competence* require courts to recognise the limits of their expertise and to be careful not to second-guess decisions reached by more competent bodies equipped with greater capacity, experience, and information. This consideration can, in many cases, also provide a reason for more intense scrutiny where courts are better equipped to decide matters than other branches of state.¹⁵⁵ The court will also be well-placed to scrutinise measures more closely where the other organs of state have failed to act or have displayed clear incompetence.¹⁵⁶

These principles are flexible and open-ended and must be weighed up in each case. What is needed is for the Court to justify openly its chosen intensity of review in each case with reference to these principles. This is not only necessary to ensure greater openness and accountability in these decisions. It is also required to give lower courts guidance in how to navigate these issues.

¹⁵² See the judgment of Cameron et al in *Barnard* (note 2 above) at paras 152–155 on the need for an intersectional approach to affirmative action. In its unfair discrimination jurisprudence, the Constitutional Court has consistently recognised the need for an intersectional approach. See further *National Coalition for Gay and Lesbian Equality and Another v Minister of Justice and Others* [1998] ZACC 15, 1999 (1) SA 6 (CC), 1998 (12) BCLR 1517 (CC) at para 113; *Bhe and Others v Magistrate, Khayelitsha and Others* [2004] ZACC 17, 2005 (1) SA 580 (CC), 2005 (1) BCLR 1 (CC) and *Hassam v Jacobs NO and Others* [2009] ZACC 19, 2009 (5) SA 572 (CC), 2009 (11) BCLR 1148 (CC).

¹⁵³ See further JH Ely *Democracy and Distrust: A Theory of Judicial Review* (1980) chs 4–6; King (note 150 above) ch 6.

¹⁵⁴ E Mureinik ‘A Bridge to Where? Introducing the Interim Bill of Rights’ (1994) 10 *South African Journal on Human Rights* 31, 32; McLean (note 148 above) 88, 211; S Fredman ‘Adjudication as Accountability’ in N Bamforth & P Leyland (eds) *Accountability in the Contemporary Constitution* (2013) 106, 108.

¹⁵⁵ On the intricacies of claims of competence, see further King (note 150 above) chs 7–9, who more accurately subdivides the broad question of competence into issues of polycentricity, expertise, and flexibility.

¹⁵⁶ See further Chan (note 18 above) who argues that the government’s claims to have greater competence should be proved with evidence, rather than being accepted on face value.

3 *Respecting the Baseline*

A fully developed doctrine of deference not only contains the principles just outlined, but also includes more determinate norms. These norms set a basic minimum intensity of review. In her writing on proportionality analysis, Cora Chan has labelled these norms the ‘baseline intensity’ of review.¹⁵⁷ A court can scrutinise laws and actions more intensely than the baseline requires, but it cannot sink below this basic minimum intensity. Beyond the baseline, the principles of deference provide guidance on when and how to vary the intensity of review. A doctrine of deference for affirmative action cases should therefore include both a baseline intensity of review and a set of principles to govern variability beyond the baseline.

It is beyond the scope of this article to provide a full account of the baseline requirements that apply in affirmative action cases. However, a few baseline norms are readily apparent. The most essential is that the *Van Heerden* test requires a proportionality analysis. For the reasons I have explained above, it would not be permissible for a Court to vary this standard of review so far that it merely applies a rationality analysis in disguise. The EEA adds further baseline requirements in assessing affirmative action, such as a prohibition on quotas and absolute job reservations. These are minimal, baseline requirements that a court cannot avoid. As explained above, the mere fact that an affirmative action measure is not a quota does not mean that it is valid. It may still fall foul of the *Van Heerden* test.

D Applying the Doctrine of Deference to *Barnard*

It is helpful to analyse how the doctrine of deference could have been applied in *Barnard* in justifying the chosen intensity of review. The majority judgment and the concurrences were clearly motivated by these considerations, but there was no real attempt to use these considerations to explain and justify why the Court’s chosen intensity of review was appropriate.

As indicated earlier, the most important question for all of the judgments in *Barnard* was how to allocate the burden of proof, including the burden of non-persuasion and the evidential and argumentative burden. As I previously explained, in allocating the burden of non-persuasion, the Court is deciding on who should bear the risk of uncertainty about the merits of the affirmative action programme. In affirmative action cases, this burden is generally placed on the persons challenging the affirmative action measure, but it may be shifted in appropriate circumstances. There is no strict, baseline norm requiring that this burden must always fall on one party. This is in contrast with the test for unfair discrimination, as the presumption of unfairness creates a firm baseline norm that the party accused of unfair discrimination will bear the burden of proving that discrimination is fair. In the absence of a baseline norm in affirmative action cases, the burden of non-persuasion will largely be determined by the three principles of deference identified above.

¹⁵⁷ Chan (note 1 above). On the baseline intensity of review under the *Harksen* test for unfair discrimination, see McConnachie (note 37 above).

The first principle, the interests at stake, provided a strong justification for placing the risk of uncertainty on Barnard. At the general level, affirmative action measures must be seen as an important component of achieving equality, in line with s 9(2) of the Constitution. The drafters of the Constitution expressed a general policy preference for affirmative action which courts cannot ignore or dismiss. Principles of democratic legitimacy therefore require courts to avoid second-guessing that general preference.

While the Constitution expresses a general preference for affirmative action, that does not mean that burden of non-persuasion will always be on the claimant. If the claimant can show that the affirmative action measure or its implementation imperils important interests, then the Court may be inclined to shift the risk of non-persuasion to the opposing party. Barnard failed to make out such a case. Had she put up some credible evidence that the implementation of the employment equity plan had a negative impact on gender equality or on service delivery, then the Court may have been more inclined to shift the burden to the National Commissioner. However, in the absence of any compelling interests at stake, Barnard rightly bore the risk of uncertainty.

Issues of institutional competence also have a bearing on this burden. A court will generally assume that senior decision-makers such as the National Commissioner are skilled and competent professionals. Absent any proof to the contrary, they are rightly slow to intervene. As the judgments indicated, the National Commissioner's failure to give evidence and further justification for the decision did complicate matters. Competence is, after all, best demonstrated by the quality of the justification provided for the decision. Nevertheless, in the absence of evidence to the contrary, the National Commissioner's competence was not called into question.

The National Commissioner certainly had an evidential and argumentative burden to justify and expand upon his cursory reasons for refusing to promote Barnard. As Cameron et al noted, accountability requires public officials to be open about their reasoning, particularly where silence can engender bitterness and distrust.¹⁵⁸ However, the fact that the National Commissioner did not discharge his evidential and argumentative burden does not necessarily impact upon the result. Once it is accepted that Barnard bore the risk of non-persuasion, any paucity of evidence on the record was to her disadvantage.

There is some uneasiness about letting the National Commissioner off the hook for his failure to provide a full justification for his actions. However, this uneasiness can be resolved if we acknowledge two things. First, the variable intensity of review means that the National Commissioner would not be let off so lightly if the circumstances were different. If Barnard had based her case on gender equality, then the burden of non-persuasion may well have been placed on the National Commissioner. Similarly, if it was shown that the National Commissioner's implementation of the employment equity plan excluded other designated groups, such as black women or disabled persons, then the Court would have been more likely to shift this burden. Second, it must also be recognised that the burden of non-persuasion is not a difficult burden to bear when the

¹⁵⁸ *Barnard* (note 2 above) at paras 110–111.

justification presented by the other side is scant or non-existent. It would not have been impossible for Barnard to discharge her burden, had a proper case been made out.

In sum, the principles outlined above can be used to offer a persuasive justification for the Court's chosen intensity of review in *Barnard*. Justifications like this are needed, not only for the sake of transparency, but also for the sake of offering guidance to lower courts in navigating these difficult issues.

VII CONCLUSION

Barnard was an opportunity to accomplish three essential tasks. First, clarity was needed on the application of the *Van Heerden* test to affirmative action measures under the EEA. That task was partially achieved when the Court confirmed that the *Van Heerden* test applies to affirmative action measures. However, the uncertainty over whether this test applies to the implementation of these measures leaves a critical gap in the law.

Second, the Court needed to clarify the appropriate standard of review to be applied. In this respect, *Barnard* is a step backwards as the divergent approaches are likely to produce greater uncertainty.

Third, the Court needed to clarify when and how it is appropriate to vary the intensity of review in scrutinising affirmative action measures under the *Van Heerden* test. As I have argued, flexibility in the application of the *Van Heerden* test is important, provided that clear and principled reasons are provided for applying the test more or less stringently. This flexibility is apparent in the way that the court modifies the standard of review, adjusts the evidential and argumentative burden, and takes it upon itself to consider evidence and arguments not squarely raised by the parties. Future cases will likely turn on these decisions. Indeed, in *Barnard*, the burden of proof proved to be the decisive matter in all judgments.

I have argued for three developments in the way in which the Court scrutinises affirmative action. First, the *Van Heerden* test should be applied under the EEA and other specialist legislation in considering both affirmative action measures *and* the implementation of these measures. Second, the Court must clarify that the *Van Heerden* test involves a proportionality analysis. Finally, and most significantly, the Court needs to develop and apply a doctrine of deference to justify variations in the intensity with which it applies the *Van Heerden* test.

This doctrine of deference has two parts. First, there are the baseline requirements, the non-negotiable minimum level of scrutiny that must be applied. Second, and more significantly in this area, there are the principles that must be weighed against each other to justify the intensity of review that is applied beyond the baseline.

Determining the appropriate intensity of review is ultimately a value-laden exercise. While these decisions manifest in the application of doctrines and formulae, they are driven by a deeper understanding of the appropriate and legitimate institutional role of the courts in adjudicating challenges to affirmative action. Unless a more transparent, principled approach is adopted here, the Court's affirmative action jurisprudence will remain obscure and uncertain.

