DE JURE

JAARGANG 46 2013 VOLUME 2

INHOUDSOPGAWE CONTENTS

Artikels/Articles

Aantekeninge/Notes

Onlangse regspraak/Recent case law

Boekaankondigings/Book Announcements

Redaksioneel

We live in a world which is becoming more and more complex. Modern technology continues to pose new questions to the law, while lawyers and governments are slow to adapt to the new challenges. This, of course, creates vast new fields of research and opens up more traditional fields to new analysis and interpretation. This volume contains a healthy dose of (at least in legal terms) newer topics which still cause much uncertainty and requires further reflection and research. There is also a good measure of traditional topics which suggests that, even after thousands of years of development and debate, the final word will probably never be said on the law relating to these matters. And nor should it. Society is not stagnant – it is in a constant state of flux and therefore the law must inevitably also be fluid.

The Austrian physicist and philosopher Ernst Mach said: "The aim of research is the discovery of the equations which subsist between the elements of phenomena". Legal research should therefore focus on law as the social equations created to address social phenomena and regulate social relations among people, whether as individuals or collectives. And because law, like society, is always changing, it remains an unfulfilled promise – an empty promise for many who lack the means to make their voices heard, or a broken promise for those whom the law does not or cannot adequately protect.

Since law can never be more than an unfulfilled promise, the notion of perfect law will forever remain an impossible dream. And imperfect law will inevitably result in countless instances where individuals or groups are let down or "slip through the cracks". Legal research should expose these flawed equations and provide the stimulus for debate on ways and means to address the shortcomings. To do this, researchers must question and challenge existing norms. They must dissect the current legal equations to expose fallacies and incorrect assumptions, to reveal flawed, invalid or outdated arguments and come up with improved equations.

Maar navorsing moet ook daarop wys dat die reg alleen nie alle sosiale probleme die hoof kan bied nie. Dit wil tans voorkom of daar wêreldwyd 'n tendens bestaan dat veral politici, maar ook ander sosiale rolspelers, die beskouing huldig dat daar kitsoplossings vir sosiale probleme geskep kan word bloot deur die reg te wysig of nuwe wetgewing uit te vaardig. Dit is 'n gevaarlike tendens wat enersyds lei tot die uitvaardiging van 'n oorvloed wetgewing wat uiteindelik onhanteerbaar word en nie doeltreffend afgedwing kan word nie. En wanneer bestaande regsreëls nie doeltreffend afgedwing word nie, verval die samelewing andersyds in 'n staat van toenemende wetteloosheid. Die feit dat minder as 20% van verkeersboetes in Suid-Afrika betaal of gevorder word en die miljarde Rande aan onbetaalde munisipale rekeninge, is skreiende voorbeelde hiervan. Die reg bied bloot die meganisme waarmee sosiale verhoudings gereël kan word, maar geen oorvloed van regsreëls kan hoegenaamd 'n

verskil maak sonder dat daardie regsreëls doeltreffend toegepas en afgedwing word nie.

Prof Steve Cornelius
Redakteur

Plain packaging and its impact on trademark law*

Louis Harms
BA LLB LLD
Adams & Adams Professor in Intellectual Property Law and Extraordinary Professor, Department of Private Law, University of Pretoria
Former Deputy President of the Supreme Court of Appeal

OPSOMMING

Ongetooide Verpakking en die Impak Daarvan op die Reg Insake Handelsmerke

Handelsmerke word ruim beskerm en wetgewers laat die eienaars van handelsmerke tot 'n groot mate toe om hulle handelsmerke ongestoord te gebruik. Daar is egter 'n nuwe wetgewende tendens op die horison wat 'n bedreiging inhou vir handelsmerke. Hierdie wetgewing, wat hoofsaaklik op tabakprodukte gemik is, vereis dat sodanige produkte slegs in voorgeskrewe ongetooide verpakking te koop aangebied mag word. Die verpakking is effekleurig en slegs die handelsnaam mag op die voorkant verskyn. Dit is die gevolg van riglyne uitgereik deur die Wêreld Gesondheidsorganisasie na aanleiding van die konvensie oor die beheer van tabakprodukte. Australië is die eerste land wat by wyse van wetgewing aan die riglyne uitvoering gegee het en hierdie wetgewing het toetsing deur die howe deurstaan. Dit wil voorkom of Suid-Afrika ook afstuur op soortgelyke wetgewing. Die gebruik van ongetooide verpakking kan egter onverwagse gevolge inhou. Vervalsing en sluikhandel word tot 'n groot mate bekamp deur die eienaars van handelsmerke wat hulle handelmerke teen misbruik beskerm. Indien tabakprodukte in ongetooide verpakking verkoop moet word, kan dit vervalsing en sluikhandel bevorder deurdat dit makliker sal wees om die verpakking na te boots en moeiliker sal wees vir die eienaars van handelsmerke om hulle regte af te dwing.

1 Introduction

Trademark owners are constantly testing the limits of their rights.[1] Courts, in general, tend to be generous in protecting trademarks and legislatures incline to leave rights holders alone. Patentees, on the other hand, are under threat like rhinoceroses, everyone wishing to get a hand on the mystical value of horns and inventions. The persistent threat to monopoly rights comes not only from legislatures and courts but more particularly from sometimes myopic vocal pressure groups also known as civil society.

---

* Prestige lecture presented at the University of Pretoria on 2013-03-13. A word of appreciation to Prof David Vaver, Oxford University, for his editorial advice and to Ms Tracy Rengecas at the Centre for Intellectual Property Law for her research assistance.

1 http://www.trademarkia.com/opposition/opposition-brand.aspx (accessed 2013-03-01).

A notable exception to this generalisation about trademarks was the Laugh-it-Off case which dealt with dilution of trademarks.[2] This is not the occasion to revisit that judgment. Others and I have done so before.

On the horizon looms a potential legislative threat to trademark law in the form of plain packaging legislation. This type of legislation typically requires that tobacco products be sold in prescribed packaging of this kind.

As appears from the illustration, the trademark may only be in the font and bland colour as indicated by the word "brand". Such limitation, one may safely assume, will be introduced through either special legislation or an amendment to health-related laws, and not by means of an amendment to the Trademarks Act.[3] It will begin with tobacco products but there is a real likelihood that it will spread to other products.

Although trademarks may exist without registration, this paper is only concerned with the larger and typically more valuable category of trademarks that are registered.

2 History

I would like to begin with some history:[4] When Christopher Columbus "discovered" the Bahamas in 1482 the local Caribs presented him with some tobacco leaves, presumably for smoking a peace pipe. Whether he had health concerns history does not tell us but as soon as he left the island he threw them, i.e. the leaves, overboard. His anti-smoking campaign was not a success.

Sir Walter Raleigh, a favourite of Queen Elizabeth, allegedly popularised the use of tobacco in England. He even received a royal patent for forming a colony in the New World. He called it Virginia, whether after his allegedly virgin queen or after Virginia leaf, the tobacco, I do not know. He was, unfortunately, not a favourite of her successor, King James 1, who had him sentenced to death on two occasions. The first sentence was commuted but the second duly executed in 1618 as a favour to the Spanish throne.

2 *Laugh It Off Promotions CC v South African Breweries* 2006 1 SA 144 (CC).
3 194 of 1993.
4 *Lapham's Quarterly* Winter 2013.

Sir Walter took a pipe of tobacco shortly before he went to the scaffold. According to John Aubrey[5] some "formal" people were scandalised but Aubrey thought that it was well and properly done to settle his spirits. Raleigh also left a small tobacco pouch in his prison cell engraved with an inscription: *Comes meus fuit in illo miserrimo tempore* ("It was my companion at that most miserable time").

There may have been another reason why King James wished him condemned: his contempt of the King's pamphlet entitled *Counterblaste to Tobacco* (1604). Royalty at the time was more literate than some of their contemporaries today. Henry VIII, for instance, wrote serious anti-reformation tracts before his anti-papal conversion. Reverting to James, as is apparent from the title, he abhorred tobacco smoking, chewing or snuffing. Its use embodied three sins: lust; intoxication; and "the greatest sin of all": that his subjects would spend their money on tobacco instead, as God had intended, on him.[6] (Chopping off someone's head as a favour to another he apparently did not regard as sinful.)

The King concluded by referring to the use of tobacco as "a custom loathsome to the eye, hateful to the nose, harmful to the brain, dangerous to the lungs, and in the black stinking fume thereof, nearest resembling the horrible Stygian (hellish) smoke of the pit that is bottomless."

To add injury to insult the King introduced what we now call a sin tax by slapping a 4,000 per cent increase on the import duty of tobacco.

3 Tobacco Wars

One has to admire the King for his insight: he realised that tobacco was dangerous to health. What he did not know was that the tobacco family of plants is of huge scientific importance as illustrated by the fact that the CSIR is developing a drug for the treatment of rabies from one of them. Whether the American-Indian tribes will have a claim for remuneration is not clear.

The health risks of the normal use of tobacco were only taken seriously during the 1970s when the term, "War on Cancer", came into vogue. However, a series of promotional cards in the 1930s included a card that explained how the latest cutting edge technology could help win the "War on Cancer."[7]

The card was in English and in Afrikaans, which tells us something about its provenance. Of special interest and ironically is the fact that it was a cigarette card issued by Max Cigarettes.

---

5 Clark (ed) *Brief Lives: Chiefly of Contemporaries, Set Down by John Aubrey, Between the Years 1669 & 1696* 189.

6 "[W]ho are created and ordained by God to bestow both their persons and goods for the maintenance both of the honour and safety of their king and Commonwealth, should disable themselves in both." (Quoted at fn 6.)

7 "The Future's War on Cancer": Smithsonian.com (accessed 2011-12-29).

Max was a well-known brand until the mid-1950s in South Africa, produced by International Tobacco Company (ITC). Its main competitor and the dominant cigarette manufacturer at the time was United Tobacco Company (UTC). UTC's competing brands in the same price class included Springbok.

UTC began a campaign to discourage the Black population from smoking Max. It employed propagandists to spread rumours about Max. Relevant for present purposes were the allegations that Max was the wrong cigarette for Blacks to smoke; that it caused coughing; that it caused tuberculosis; and (if my childhood memory serves me right) that Max caused cancer.

In the process UTC systematically destroyed the Max mark. Litigation followed.[8] The court found that the statements were false, that they were made with knowledge of their falsity, and were made maliciously. An award of damages of £580,800, which translates into some R42,000,000 at present day values, was made. UTC did not prosecute an appeal.


THIS AGE OF
POWER & WONDER
A SERIES OF 250 PICTURES
No. 6. WAR ON CANCER.

When scientists first began to create synthetic radio-activity, to make substitutes for radium, by bombarding certain atoms with millions of electron-volts, someone suggested "Why make radium to cure cancer? Use the bombarding atoms direct". This suggestion was adopted by the use of very high voltage X-rays. Many successful experiments have been made.

No. 6. OORLOG TEEN KANKER.

ISSUED WITH
MAX CIGARETTES

If one would wonder why the court did not find that the cigarettes did have deleterious health effects the answer must be that UTC could hardly, without destroying its own business, seek to justify the allegations. Instead of relying on justification its case was one of general denial.

UTC in due course had its comeuppance through Rembrandt and Rembrandt's brilliant use of trademarks.

4 Framework Convention on Tobacco Control

The World Health Organisation Framework Convention on Tobacco Control (WHO FCTC)[9] is, according to the WHO, a treaty that reaffirms the right of all people to the highest standard of health and represents a paradigm shift in developing a regulatory strategy to address addictive substances. In contrast to drug control treaties this treaty focuses on demand and supply reduction strategies, and not on prohibition.

8 *International Tobacco Co v United Tobacco Co* 1955 2 SA 1 (W).
9 WHO website http://www.who.int/fctc/text_download/en/index.html (accessed 2013-03-01).

The treaty, says the WHO, was developed in response to the globalisation of the tobacco epidemic, which is facilitated through trade liberalisation, direct foreign investment, global marketing, transnational tobacco advertising, promotion and sponsorship, and international movement of counterfeit cigarettes.

The treaty was signed by 169 countries which meant that they would strive in good faith to ratify, accept, or approve it, and show political commitment not to undermine its objectives. Notable non-signatories include the USA and Switzerland.

The treaty itself does not deal with trademarks. Of relevance for this discussion is Article 11, which requires of countries to adopt effective measures to ensure:

> (i) that tobacco product packaging and labelling do not promote a tobacco product by any means that are false, misleading, deceptive or likely to create an erroneous impression about its characteristics, health effects, hazards or emissions; and
>
> (ii) that any outside packaging and labelling should carry health warnings that should be 50 per cent or more of the principal display areas.

And then there are Guidelines. These provide that countries "should consider [I emphasise *should consider*] adopting measures to restrict or prohibit the use of logos, colours, brand images or promotional information on packaging other than brand names and product names displayed in a standard colour and font style."

The Guidelines add that the effect of advertising or promotion on packaging can [I emphasise *can*] be eliminated by requiring plain packaging: black and white or two other contrasting colours; nothing other than a brand name, a product name and/or manufacturer's name; and no logos.

The motivation for these Guidelines, it is said, is to:

> [i]ncrease the noticeability and effectiveness of health warnings and messages, [to] prevent the package from detracting attention from them, and [to] address industry package design techniques that may suggest that some products are less harmful than others.

As could be expected, the treaty came in for heavy criticism from not only those with direct interests in the tobacco industry but also those on the periphery such as intellectual property lawyers and other industries that depend on trademarks. One may immediately think of the liquor industry. And one could also envisage a Coca-Cola bottle, 70 per cent covered with a picture of someone with rotting teeth. Interestingly, the Coca-Cola Company was put on trial in the USA during 1909 for selling an adulterated and deleterious product. The issue was not the fact that

the product contained small quantities of cocaine or that it had large quantities of sugar, but because it contained added caffeine.[10]

One could also picture a McDonald's hamburger cover with a dead cow or an obese kid.[11] Or a Toyota covered with pictures of blood-spattered victims of taxi accidents.

A German newspaper facetiously suggested that every male organ should carry both a health warning and one about the costs of raising an unexpected child.

The legal objections to these initiatives fall into three broad classes: (1) they are in conflict with international law; (2) they are unconstitutional; and in any event (3) they undermine basic principles of trademark law.[12] For the sake of convenience I shall address the third issue first.

5 Trademark Rights

Before dealing with the rights implicated, the following has to be stressed for the sake of context. Although trademarks are intangible property, there is no constitutional right to a trademark. The rights arise through registration and although we prefer to call trademarks intellectual property they are, as we learnt already in 1883, more properly referred to as industrial property.

Like all rights, a trademark right is not absolute but relative, and may in any event be trumped by other rights.

A trademark is said to be a negative right. It is a right to prevent others from using the same or a confusingly similar trademark for the same or similar goods or services. All things being equal, its ownership gives a preferential right to use to the owner but not an absolute right to use the mark on the particular goods for which it is registered. To illustrate, simply because one holds a trademark for a prohibited substance does not mean that one is entitled to market that substance, with or without the mark.

The principles of trademark law affected by plain packaging regulation are the following:

(i) Trademarks serve as badges of origin to prevent confusion or deception as to the origin of the product.

10 *United States v Forty Barrels and Twenty Kegs of Coca Cola, the Coca Cola Company of Atlanta, Georgia* 241 US 265.

11 "In Fight Against Obesity, Drink Sizes Matter" http://well.blogs.nytimes.com/2012/10/22/in-fight-against-obesity-drink-sizes-matter/ (accessed 2013-03-01).

12 Halabi "International trademark protection and global public health: A just compensation regime for expropriations and regulatory takings" 2011 *Cath U LR* 325.

(ii) They also have other functions, such as an advertising function by acting as silent salesmen conveying psychological messages about the merit of the product and they thereby implicitly guarantee quality.

(iii) The use principle: In order for marks to be registrable, the applicant must have the intention of using them; and if they are not used, they may be removed from the register.

There can be no doubt that plain packaging regulations based on the Guidelines make it impossible to use figurative trademarks. Well-known are for instance the camel of Camel and the red rooftop mark of Marlboro.[13] These can, consequently, not be used as badges of origin.

Concerning name marks, the trademark owner is under the Trademarks Act[14] entitled to use them in any manner or form or colour but regulation will limit or destroy that right. And as to confusion or deception, a mark risks losing its distinctiveness if it cannot be represented without restriction; and the less its distinctiveness the less likely it will be able to reduce or prevent confusion. Regulations requiring the same lettering, the same colour etcetera, will no doubt impinge on the distinctiveness of a trademark.

The other functions of trademarks are not protected by statute except through dilution (to the extent that it still exists) but dilution through edict such as plain packaging regulation is not actionable under the Trademarks Act.[15]

That brings me to a fundamental principle of trademark law, namely use: as long as the applicant has the intention to use a mark, it is registrable. It is irrelevant whether he might be prevented from using it. And as for being struck off the register, the Act has an inbuilt protection: a trademark may not be removed on the ground of non-use if that was due to special circumstances in the trade and not to any intention not to use or to abandon the trademark.

In introducing its plain packaging legislation, Australia anticipated an argument that the statute would lead to the invalidation of trademarks. The Australian statute (to which I shall return) accordingly contains savings provisions: The registrability of trademarks is not to be prejudiced by the constraints on use; and they are also not to be deprived of registrability or revoked for non-use, or because their use in relation to tobacco products would be contrary to law. These saving provisions, I believe, are inherent in our Trademarks Act.[16]

13 *Phillip Morris Products v Marlboro Canada* 2010 FC 109; *Marlboro Canada Limited v Philip Morris Products SA* 2012 FCA 201.
14 194 of 1993.
15 194 of 1993.
16 194 of 1993.

What this means is that trademark law does not have a number of immutable principles and that an act that imposes limitations on trademark use simply amends the Trademarks Act[17] *pro tanto* in relation to the particular goods or services.[18] Any legal attack must consequently be based on either international law (which has been internalised) or on constitutional law.

6 International Law

It is easy to make short shrift of the reliance on international law. There are two main international conventions regulating trademark law: the Paris Convention on Industrial Property of 1883 and the Agreement on Trade Related Aspects of Intellectual Property Rights (TRIPS) (which is part of GATT) of 1994. South Africa is a member of Paris and, by virtue of its membership of the WTO, bound by TRIPS.

The Paris Convention does not deal much with the substantive rights of trademark owners and is of no assistance.

The TRIPS agreement deals, albeit to a very limited extent, with substantive IP rights.[19] However, TRIPS (and the Paris Convention) does not give rights to individuals or corporations but is only binding as between party states. It is accordingly not enforceable under national law, which means that the issue can only be raised by a tobacco exporting country against a tobacco importing country.

The issue of plain packaging has already been discussed in the TRIPS Council in 2011, and in July 2012, the Dominican Republic notified the WTO that it had launched a dispute settlement case against Australia's plain packaging law and on 9 November 2012 it requested the establishment of a WTO dispute settlement panel.[20] To illustrate what upholding a complaint means, one may refer to the spat between Antigua and the USA. The WTO upheld Antigua's complaint about the ban on internet gambling by the USA and the USA failed to rectify the position. This entitles Antigua to retaliate for the $21m loss its online gaming industry is said to have suffered. It has given notice of its intention to ignore US intellectual property rights, including trademarks, to that extent.[21]

17 194 of 1993.
18 Bonadio "Plain packaging of tobacco products under EU intellectual property law" 2012 *Eur IP R* 599.
19 Mitchell & Voon "Implications of WTO Law for Plain Packaging of Tobacco Products" http://ssrn.com/abstract = 1874593 (accessed 2013-03-01).
20 Saez "LDCs to Press for Extension for TRIPS, Plain Packaging Back" http://www.ip-watch.org/2013/02/26/wto-ldcs-to-press-for-extension-for-trips-plain-packaging-back/ (accessed 2013-02-26).
21 http://nationalinterest.org/commentary/intellectual-property-pirates-the-caribbean-8136 (accessed 2013-03-01).

I do not rate the chances of success of the Dominican Republic very high. TRIPS regards a trademark right as negative right: it is the right to prevent others from using the trademark under certain circumstances. TRIPS considers the function of trademarks to be that of a badge of origin and it does not purport to deal with its subsidiary functions, such as that of silent salesman.

Although TRIPS states that the nature of the goods or services to which a trademark is to be applied may not form an obstacle to its registration, it does not say that one is thereby entitled to use such a trademark.

The only other international remedy, which is in the hands of an investor, may be found in bi-lateral trade agreements but that is country specific and a subject beyond the scope of this paper.[22]

7 Australia is no Longer Marlboro Country: Plain Packaging in Australia

Australia, in spite of strenuous industry opposition, was the first country to act on the Guidelines. Although its adoption of the Guidelines was said to have been done in order to comply with its international obligations that was simply posturing. There was no international obligation.

It is not necessary for purposes of this presentation to set out the detail of the Australian provisions. It suffices to state that they follow the Guidelines conscientiously and the pictorial examples illustrate their effect.[23]

The Commonwealth of Australia is a federation and its ability to make laws is circumscribed by its Constitution. An attack on the validity of legislation has, therefore, to be based on the allegation that the particular law is not authorised by the Constitution.

The tobacco industry sought to have the plain packaging legislation declared unconstitutional. Although, according to it, a number of IPRs were involved, I shall limit the discussion to the trademark aspect of the case.

The Australian Constitution confers upon the Commonwealth Parliament the power to make laws with respect to "[t]he acquisition of

22 Vadi "Global health governance at a crossroads: Trademark protection v tobacco control in international investment law" 2012 *Stan J Int L* 93.

23 http://www.yourhealth.gov.au/internet/yourhealth/publishing.nsf/content/ictstpa#.UD13jNVwVjR (accessed 2013-03-01).

property on just terms ... for any purpose in respect of which the Parliament has power to make laws".[24]

The matter in due course reached the High Court of Australia.[25] The essential questions for decision were, first, whether IPRs are "property" and, second, whether the limitation on the use of trademarks amounted to "the acquisition of property".

It would appear that the court was fairly unanimous in holding that trademarks are property within the meaning of the term in the section of the Constitution and entitled to constitutional protection in spite of the fact that trademark rights are negative in nature.

As to the second question the majority (6 to 1) held, I think, that trademark registration does not confer a liberty to use a trademark free from restraints found in other statutes or the general law. And although trademark rights are in substance, if not in form, denuded of their value and thus of their utility by the plain packaging provisions that does not amount to an "acquisition": rights of property may be extinguished without being "acquired".

As French CJ explained (at para 42):

> Taking involves deprivation of property seen from the perspective of its owner. Acquisition involves receipt of something seen from the perspective of the acquirer. Acquisition is therefore not made out by mere extinguishment of rights.

The Commonwealth, by imposing limitations, did not acquire anything.

And he concluded as follows (at para 44):

> In summary, the TPP Act is part of a legislative scheme which places controls on the way in which tobacco products can be marketed. While the imposition of those controls may be said to constitute a taking in the sense that the plaintiffs' enjoyment of their intellectual property rights and related rights is restricted, the corresponding imposition of controls on the packaging and presentation of tobacco products does not involve the accrual of a benefit of a proprietary character to the Commonwealth which would constitute an acquisition.

Matthew Rimmer, quite clearly a passionate anti-smoking fanatic, enthused that "the High Court's ruling is one of the great constitutional cases of our age."[26] I would have thought, considering the issue before the court and the wording of their Constitution, that the result was fairly predictable, if not inevitable.

---

24 It is the subject of the Australian film *The Castle* (1997).

25 *JT International SA v Commonwealth of Australia; British American Tobacco Australasia Limited v The Commonwealth* [2012] HCA 43.

26 http://theconversation.edu.au/the-high-court-and-the-marlboro-man-the-plain-packaging-decision-10014 (accessed 2013-03-01).

Rimmer added that "the ruling will resonate throughout the world – as other countries will undoubtedly seek to emulate Australia's plain packaging regime." That no doubt is correct and once again fairly predictable and early copycats predictably will be the European Union[27] and South Africa. New Zealand has already given notice of its intention to adopt plain packaging legislation.

8 Constitutional Arguments

A law is unconstitutional if it impinges on an entrenched right but it will nevertheless be saved to the extent that the limitation is justifiable.

It is fair to assume that, as in Australia, any attack on plain packaging legislation will primarily be based on the property clause in our Constitution. Not being immersed in constitutional law concepts I shall keep it simple.[28]

According to section 25(1), no one may be deprived of property except in terms of a law of general application; and no law may permit the arbitrary deprivation of property. This provision will be of no assistance because a plain packaging law does not deprive the trademark owner of any trademark right but only regulates or limits the exercise of that right.

Section 25(2) provides that property may be expropriated only in terms of law of general application for a public purpose or in the public interest; and subject to compensation. The question then is whether a plain packaging law would amount to expropriation of property. According to our law trademarks are property, which leaves for consideration the meaning of "expropriation". Our courts interpret the word as requiring not only dispossession or deprivation but also appropriation by the expropriator (on behalf of itself or another party) of the particular right. It should immediately be obvious that although we are dealing with a provision that differs in its wording from the Australian, the result will inevitably be the same.

Assuming that I am wrong or that other provisions of the Bill of Rights might be implicated, I do believe that the limitation will be found to be justifiable. In *Prince v President of the Law Society*,[29] a Rastafarian argued that the prohibition of the use of cannabis for religious purposes infringed his right to freedom of religion. The Constitutional Court in a split decision found that although the prohibition infringed his right the limitation nevertheless was justified on general health grounds. This, to me, is a clear indication that the Court will find that any limitation, direct or indirect, on the use of tobacco would be justifiable. The majority said:

27 During September 2013, the European Parliament adopted a somewhat diluted version.

28 For a fuller discussion of expropriation principles in another context see Van der Vyver "Nationalisation of mineral rights in South Africa" 2012 *De Iure* 125.

29 2002 2 SA 794 (CC).

> In a democratic society the legislature has the power and, where appropriate, the duty to enact legislation prohibiting conduct considered by it to be anti-social and, where necessary, to enforce that prohibition by criminal sanctions. In doing so it must act consistently with the Constitution, but if it does that, courts must enforce the laws whether they agree with them or not.
>
> The question before us, therefore, is not whether we agree with the law prohibiting the possession and use of cannabis. Our views in that regard are irrelevant. The only question is whether the law is inconsistent with the Constitution.

Another signpost is the refusal of the Constitutional Court to grant leave to appeal in *British American Tobacco South Africa (Pty) Ltd v Minister of Health.*[30] The case concerned the constitutionality of a provision which states that "no person shall advertise or promote … a tobacco product through any direct or indirect means, including through sponsorship of any organisation, event, service, physical establishment, programme, project, bursary, scholarship or any other method." The SCA held that although the provision infringed some constitutional rights (particularly that of free speech) it was, from a public health perspective, justifiable. The Constitutional Court refused to reconsider this conclusion.

9 Pampers for South Africa?

The Hungarian author, Sándor Márai, dealt with smoking addiction in this monologue:[31]

> How do you get on with the tobacco habit? It's a struggle, isn't it? I couldn't go on – not with smoking but with the struggle. There will be a day when that too has to be faced. One adds up the facts and decides whether to live five or ten years longer by not smoking, or to surrender to this petty, shameful passion that no doubt kills but, until it does so, offers you a peculiar calming yet exciting experience. After fifty years, it becomes one of life's major questions. The answer to that question was angina and the decision to carry on exactly as before until I die. I'll not stop poisoning myself with this bitter weed because it's not worth it. You say it's not so difficult to give up? Of course it's not that difficult. I've done it before, more than once, while it was worth it. The trouble was, I'd spend the whole day not smoking. That's something else I'll have to face one day. People should resign themselves to certain weaknesses, to their need for a soporific of some sort, and be prepared to pay the price. It's so much simpler that way. Yes, but then they say: you should have more courage. My answer to them is: I may not be the bravest of men, but I am courageous enough to live with my desires.

Courageous words but we have to accept that the State has assumed the right or obligation to decide which desires are acceptable and that one should live one's life according to its dictates. In that regard the doctrine of voluntary assumption of risk is dead. One could simply refer to the history of opium without suggesting that opium and tobacco are similar.

30 [2012] 3 All SA 593 (SCA).
31 Márai *Portraits of a Marriage* (translation by Szirtes (2011)) 133.

The Chinese Emperor sought to stamp out its use in 1838. However, British traders had an interest in importing the stuff into China. They used it for barter to save payment in silver. The actions of the Emperor led to the Opium War, which China lost and the opium trade formed the basis of the fortunes of what is today a very respectable company in the Far East, Jardine-Matheson.

The use of opium for recreational purposes was socially acceptable in Europe as evidenced by the *Confessions of an English Opium-Eater* (1821), an autobiographical account by Thomas De Quincey, about his laudanum addiction and its effect on his life. (Laudanum is an alcohol based drug containing opium.) Those of you who have seen the recent film, *Die Wonderwerker*, on the life of Eugene Marais will recall that his supplier of opium warned him that General Louis Botha, prime minister until 1919, intended to prohibit the importation of opium, the importation having been legal until then.

De Quincey's work could be compared to another classic, that of Italo Svevo: *Zeno's Conscience*. It is a brilliant novel, described hyperbolically as the greatest comic novel of the twentieth century, partly about a man's attempt to give up smoking. It is in that respect autobiographical: the hero, like the author, was always smoking his last cigarette. After publication and in real life the author sustained serious injuries during a car accident. Probably realising that he was close to death's door he asked for a cigarette, which he said would *really* be his last. For health reasons the doctor, a bit of a nanny, refused and Svevo was dead within a few hours. There is a moral somewhere but I am not sure what it is.

The choice for governments appears to be between restriction and prohibition. Prohibition, of liquor or underage sex, does not work but I do not think that this is the reason why government does not prohibit the use of tobacco.

Without being too cynical, the reason is because government does not wish to forego the resultant income.[32] Our government, it would appear, earned R32 billion per annum on tobacco excise and related taxes before the increase announced on 27 February 2013.

The tension between health and state income was also underlined when the relevant health departments decided to take on advertising in the liquor industry. The Department of Trade and Industry (DTI) responded by stating that such a decision could not be taken without considering the economic impact.[33] Since writing this paper Cabinet nevertheless has approved such a bill for consideration by Parliament.

---

32 The case for banning is set out by Biegler at http://theconversation.edu.au/why-banning-cigarettes-is-the-next-step-in-tobacco-control-8915 (accessed 2013-03-01).

33 http://www.bdlive.co.za/business/media/2013/02/22/alcohol-advertising-ban-not-feasible (accessed 2013-03-01).

It is significant is that our government, according to estimates, loses R8 billion per annum due to smuggling of cigarettes.[34] The estimate for the United Kingdom is £3 billion. According to statistics released by the European Union, cigarettes and tobacco products are number 1 or 2 on the list of counterfeit products.

The leader behind the attack on the oil facility in January 2013 in Algeria was one Belmokhtar. He, it is alleged, generated millions of dollars through the smuggling of tobacco earning him the nickname "Mr Marlboro".[35]

There is logic behind this. Smugglers do not pay the notoriously high sin excise taxes and on that score alone make huge profits. And they run low sentencing risks compared to drug smuggling.

South Africa during February 2013 entered into a convention to contain the smuggling of tobacco products, presumably with an eye on the R8 billion. Apart from our customs and tax laws we have the Counterfeit Goods Act,[36] all supposed to confine smuggling. How another convention will make a difference is difficult to comprehend.

Why do I raise all this while dealing with plain packaging? Let me begin with an analogy. It is common knowledge that Prohibition in the 1920s United States consolidated the hold of large-scale organised crime over the illegal alcohol industry and increased its other activities. Similarly, the War on Drugs, intended to suppress the illegal drug trade, instead consolidated the profitability of drug cartels.

Those at the forefront of the fight against smuggling are trademark owners. They lose more than governments. Their best weapon is a trademark. A diluted trademark is of little value and can easily be counterfeited. Trademark owners will no longer have the ability or much motivation to contain smuggling. In other words, plain packaging will lead to more counterfeiting and smuggling – and greater loss to the fiscus.

That is known as the law of unintended consequences.[37]

34 http://www.fin24.com/Economy/Smuggled-cigarettes-costs-R12bn-in-taxes-20121105 (accessed 2013-03-01).

35 http://www.nytimes.com/2013/01/20/world/africa/in-chaos-in-north-africa-a-grim-side-of-arab-spring.html?pagewanted = all&_r = 0 (accessed 2013-03-01). He was also known as One-Eye and was allegedly killed by the forces of Chad on 2013-03-02.

36 37 of 1997.

37 http://en.wikipedia.org/wiki/Unintended_consequences (accessed 2013-03-01).

Collective sale of sports television rights in the European Union: competition law aspects

Ian Blackshaw
LLM LLD
Visiting Fellow, Centre for Intellectual Property Law, University of Pretoria, Fellow, TMC Asser International Sports Law Centre, Solicitor of the Supreme Court of England and Wales, Visiting Professor at several UK and European Universities

OPSOMMING

Kollektiewe Verkope van Televisieregte vir Sport in die Europese Unie: Mededingingsfasette

Televisieregte vir sport word wêreldwyd vir astronomiese bedrae verhandel. In die Europese Unie (EU), word hierdie regte gewoonlik deur sportliggame en -ligas op 'n kollektiewe rondslag verhandel, ten einde die waarde daarvan te verhoog. Sulke reëlings is egter onderhewig aan EU Mededingingsmaatreëls, wat daarop gerig is, nie om mededinging te verhoed nie, maar om onregmatige mededinging te verbied en die spreekwoordelike speelveld gelyk te maak. Hierdie artikel oorweeg die maatstawwe, soos neergelê deur die mededingingsdirektoraat van die Europese Kommissie en uitgewys in regspraak van die Europese geregshof, om te bepaal of sodanige kollektiewe verhandelinge van geval tot geval regmatig is. Dit oorweeg ook die beperkings wat die sogenaamde EU "*Television Without Frontiers*" *Directive* plaas op die verhandeling van TV-regte deur sportliggame, en wat vereis dat bepaalde sportgeleenthede van sosiale belang in elkeen van die 28 lidstate van die EU algemeen beskikbaar moet wees op "gratis-oor-die-lug-televisie". Die artikel fokus ook op die onlangse appèlsake wat in die Europese geregshof ten aansien van die FIFA sokker-wêreldbeker en UEFA Europese kampioenskap beslis is.

1 Introductory Remarks

At first sight, one may reasonably ask, what has sport got to do with the European Union (EU); and what has the EU got to do with sport? Any connection between the two might seem bizarre. Sport is a social and leisure activity, while the EU is essentially a single market for business, comprising some 505 million people. But apart from the social and health aspects, sport has an important economic dimension too.

In fact, sport is big business globally, representing more than 3% of world trade, and in the EU sport accounts for 3.7% of the combined GNP of the twenty-eight member states and employs 5.4% of the EU labour force. So there is much at stake both on and off the field of play in Europe, which, as far as sport is concerned globally, punches well above its size and weight!

Apart from the importance of sport in the EU from a health and social point of view, where sport involves an economic activity, EU Law in general, including the internal market rules and especially the freedom of movement of employees' provisions, and EU Competition Law in particular, including restrictive agreements and abuse of dominant positions and their "extra-territorial" effects, kicks in, as the famous – or, depending upon your point of view, infamous – Jean-Marc Bosman and subsequent related cases demonstrate.

As many sporting events take place on a European-wide basis, such as the UEFA Champions' League, for example, it is vital for all those involved in the organisation, administration, marketing and broadcasting of such events and their professional advisers to have a clear understanding of the application of a developing EU Sports Law (both "soft law" and formal decisions ("hard law")) to all of these activities. It should be noted that breach of the EU competition law can result in hefty fines of up to 10% of the world-wide turnover of a company or organisation involved in such "offences".

Although there was no specific provision in the founding European Community Treaty of 25 March 1957 giving the European Union (EU) – as it is now known – any competence in the field of sport, the EU, since the *Walrave & Koch v Union Cycliste Internationale* decision of the European Court of Justice in 1974,[1] has been ready to intervene in those cases in which "the practice of sport ... constitutes an economic activity". And it often does nowadays!

More recently, however, the signature in Rome on 29 October, 2004 of the Treaty establishing a Constitution for Europe was a significant milestone, not only for the European Union, but also for European sport. For the first time in its history, sport became an integral part of the primary law of the EU. There is now a specific legal basis for sport. The so-called Sport Article is to be found in Article 165 of the new Treaty on the Functioning of the European Union (TFEU).

Sub-paragraph 2 of paragraph 1 of Article 165 makes the following provision:

> The Union shall contribute to the promotion of European sporting issues, which taking account of the specific nature of sport, its structures based on voluntary activity and its social and educational function.

In effect, the Article recognises the social and political role and importance of sport at the European level.

Sub-paragraph 2 of the Article makes further provision as follows:

> Union action shall be aimed at:

1 Case 36/74 [1974] ECR 1405.

(i) developing the European dimension in sport, by promoting fairness and openness in sporting competitions and cooperation between bodies responsible for sports, and by protecting the physical and moral integrity of sportsmen and sportswomen, especially the youngest sportsmen and sportswomen.

Finally, sub-paragraph 3 of the Article looks at the wider relationship between sport in the Union and beyond its borders and provides as follows:

> The Union and Member States shall foster cooperation with third countries and the competent international organizations in the field of education and sport, in particular the Council of Europe.

In fact, it may be mentioned, *en passant*, that the Council of Europe has been very active over the years in the sporting arena – not least in the area of doping in sport.[2]

2 The Sale of Sports Television Rights in the European Union

The rise of sport as a global industry is largely the result over the years of the marketing of sports, sports persons, teams and events, originally in the United States of America (USA), and subsequently in Europe and elsewhere. This has led to the establishment of a world-wide discrete sports marketing industry, due to the vision and pioneering work of Mark McCormack in the USA, through his company, International Management Group; and in Europe, by Horst Dassler, of the German sports goods manufacturer Adidas, through his Swiss company International Sport Leisure and Culture, which he founded. Sadly, neither of these pioneers is alive today to see the extent to which sports marketing has grown and enjoy the full fruits of their work.

Of the sports marketing mix, which includes sports sponsorship, merchandising, endorsement of products and services, and corporate hospitality, perhaps the most important and lucrative one is the sale and exploitation of sports broadcasting rights, especially sports television rights, around the world,[3] which contribute mega sums to many sports and sports events, including the Summer and Winter Olympic Games and the FIFA World Cup. Indeed, it is fair to say that, without the mega sums generated by sports broadcasting, such major events – and, in fact, many others – could not take place and consequently sport – and sports

2 For further information on the scope of the activities of the Council of Europe in sport, see the Siekmann & Soek (eds)*The Council of Europe and Sport: Basic Documents* (2007).

3 For a comprehensive view of the importance and the scope of sports TV rights in several major sporting countries around the world, see the Blackshaw, Cornelius & Siekmann *TV Rights and Sport Legal Aspects* (2009).

fans – would be the losers.[4] In this respect, the commercialisation of sports broadcasting rights may be considered as the oxygen of sport. There is a symbiotic relationship between sport and television broadcasting. Indeed, according to Griffith-Jones[5] "[t]his marriage between sport and television is one made in heaven" and according to Parrish[6] "[t]he broadcasting sector and sport have ... revolutionised each other".

And the significance of new technology – especially broadband and quicker access to the Internet and web-streaming of sports events – in the development and financial importance of sports broadcasting rights cannot be over emphasised as Verow, Lawrence and McCormick[7] rightly point out:

> In many ways, the rise of new platforms for the dissemination of media products and the inevitable rise of sport as the global media property it now is have been intertwined. Just as the formation of the FA Premier League and the rise of satellite pay television through BSkyB seemed inextricably linked, so when new platforms, such as the proliferation of digital television channels or the exploitation for broadcast or quasi broadcast purposes of internet and mobile telephony platforms, come to the fore, their usual test bed in terms of content is in sport. It seems that only sport has the pulling power nationally and internationally to justify the sort of investments needed to bring new media platforms to market, and maybe sport is alone considered sufficiently popular for the uptake by new customers properly to reflect the potential of the medium rather count simply as a commentary on the first content offered through it.

For example, the English FA Premier League – the world's most popular and most financially successful football league – have recently sold their live broadcasting rights to their matches for the seasons 2013-2016 for a record sum of £3.018 billion. The sale of additional rights packages, comprising overseas rights, highlights and mobile phone and internet rights, is expected to increase this amount to a staggering sum of £5 billion!

Again, the lion's share of these rights has been sold to the satellite broadcaster, BSkyB, which will be shown as part of its *Sky Sports* package on a subscription basis. BSkyB is owned by the Australian media magnate, Rupert Murdoch, through his Group, News International, who, incidentally, considers "sports as a battering ram and a lead offering" in all his pay television operations around the world.[8]

Sports bodies and leagues have traditionally sold their television rights on a joint basis on behalf of their constituent members, in order to

4 See Blackshaw "Sports Marketing, Sponsorship and Ambush Marketing" in *Sports Law* (2006) (ed Gardiner).
5 Griffith-Jones *Law and the Business of Sport* (1997) 289.
6 Parrish *Sports law and policy in the European Union* (2003).
7 Verow, Lawrence & McCormick *Sports Business* (2005) 321.
8 Address at the AGM of News Corporation on 1996-10-15 in Adelaide, Australia.

leverage the value of these rights, and this practice has raised a number of competition law issues over the years, which continue to the present day.

Indeed, the collective purchase and sale of sports broadcasting rights, particularly television rights, has occupied the attention of national and EU competition law authorities for many years: for example, the EBU (European Broadcasting Union) Sports television rights; the UEFA Champions League Football Competition; the German Football Federation (DFB); and BBC/BSkyB English Premier League cases.

In 1998, A.M.Wachmeister, a member of the Competition Directorate of the European Commission at that time, issued some helpful guidelines on the application of EU Competition Rules to Sports Broadcasting. Although to some extent dated, these guidelines are still valid today and can be applied to new media situations.

Subsequently, the EU Commission investigated the arrangements for the sale of the television and commercial rights to the popular UEFA European Champions' League Football Competition. The Commission issued a Notification (IP/01/1043) and Background Note (MEMO/01/271) on 20 July 2001. The situation in this case may be summarised as follows:

The European Competition Commission was essentially concerned about the following elements of the way the television broadcast rights to the European Champions' League were being sold:

(i) Exclusivity

(ii) Single Broadcaster

(iii) On a Territory by Territory Basis

(iv) Long-term Contracts

The aim of the European Commission's intervention was "(t)o ensure that the European sports fans can benefit from a wider coverage of top European football events".[9] According to the Commission, the form of joint selling of the television rights by UEFA had "(a) highly anti-competitive effect by foreclosing television markets and ultimately limiting television coverage of those events for consumers".[10] The Commission also considered that joint selling of television rights is "(n)ot indispensable for guaranteeing solidarity among clubs participating in a football tournament".[11] And furthermore "[i]t should be possible to achieve solidarity without incurring anti-competitive effects".[12]

9 EU Commission Notification IP/01/1043 and Background Note (MEMO/01/271) of 2001-07-20.
10 *Ibid.*
11 *Ibid.*
12 *Ibid.*

Indeed, in many submissions by third parties, including broadcasters and sports rights agencies, made to the Commission, UEFA's solidarity measures were characterised as "inefficient, insufficient and conducted in a non-transparent way".[13] In other words, a sham!

The Commission's stance was that by furthering competition in the broadcasting market this will lead "[t]o better quality television coverage and lower subscription fees".[14]

Also, the Commission considered that, in joint selling arrangements, there is a reluctance to give licences to apply new technologies, such as the Internet and UMTS, because broadcasters fear that it will devalue their television rights.

The Commission, therefore, called on UEFA to submit "constructive proposals [in order] to guarantee open access to television coverage of football".[15]

After many discussions and negotiations between the parties, new selling arrangements were put forward by UEFA to the Commission, which were acceptable, and the matter was settled amicably (so-called "soft law") in December 2003. The new arrangements came into force as from the 2003/2004 football season.[16]

More recently, the EU Commission investigated the broadcast selling arrangements for the English FA Premier League (EPL) and, after protracted negotiations between the parties, extracted a number of important concessions from the EPL. These included from 2007 onwards that no single broadcaster will be able to buy all the rights of the centrally marketed live rights packages; both the EPL (in respect of all matches) and the clubs (in respect of those matches in which they participated) will have the right to provide video content on the Internet as of midnight on the night of the match; clubs will be able to provide mobile content as of midnight following the match; the EPL has increased radio broadcasting; and has also allowed two matches to be broadcast live nationally on Saturday afternoons.[17]

Mention should also be made of the German Bundesliga (DFB) case, also involving the collective sale of broadcasting rights. The same EU competition law considerations as in the UEFA and EPL cases were applied in the DFB case. These included:

(i) The unbundling of rights into separate rights packages for television broadcasting and mobile platforms;

(ii) The possibility for individual clubs to exploit certain unsold rights and rights unused by the initial purchaser;

13 *Ibid.*
14 *Ibid.*
15 *Ibid.*
16 See Case 37398, OJ 2003 L291/25.
17 See Case 38173, OJ C 7.

(iii) As well as the exploitation of deferred rights and rights for internet broadcasting and telephony broadcasting markets.[18]

(iv) Also, the rights were to be sold by way of a public tender process and exclusive contracts were not to exceed three football seasons.[19]

3 Some Further EU Competition Law Issues on the Sale of Sports Television Rights

The leading 2003 Decision of the European Commission involving the collective selling of the broadcasting rights to the UEFA European Champions' League (mentioned above),[20] which has been used as kind of "template" in subsequent sports broadcasting cases at the national level, and also the unresolved legal questions regarding the matter of the so-called "organisational solidarity" in sport[21] – considered to be legally and politically sensitive – are of crucial importance and worthy of further critical analysis and study.

Following this Decision, the Commission requires the following conditions for the sale of sports television rights to be satisfied:

(i) An open tender;

(ii) An "unbundling" of the offer allowing more than a single buyer;

(iii) No excessive exclusivity (a term of three years being regarded as a general norm); and

(iv) No automatic renewal (regarded as a disguised extension of the term of the exclusivity).[22]

The Commission's aims and critical factors in relation to opening up competition within the single EU market in the field of sports broadcasting rights may be summarised in the following remarks of the

18 With the introduction of 3G and 4G mobile phone technology allowing the delivery of sport content to smart and iphones, these rights are also in demand and have also increased in value!

19 See Case 37214, OJ 2005 L 134/46.

20 Dec 2003/778.

21 The concept of "solidarity" in sport in general and in football in particular is that the revenues generated by the sale of sports rights, especially TV rights, should "trickle down" to and be available to be used for the benefit of the so-called "grass roots" constituents and programmes of the sport concerned. This has tended so far, for example, in the case of the UEFA European Champions League to be more of a theoretical concept rather than a practical one!

22 Ungerer "Commercialising Sport: Understanding the TV Rights Debate" speech delivered in Barcelona on 2003-10-02, in which, *inter alia*, Ungerer argued that "there must be a clear separation between sports regulation and the commercialisation of sport". And added: "TV is of high significance for football clubs, 30-70% of football clubs' revenue come from TV, and this explains why sometimes our efforts [the Commission] to bring joint selling into line with Competition law requirements meet a certain anxiety – even bitterness – on the side of some leagues, and are initially misunderstood."

Competition Commissioner at the time, Neelie Kroes, in the context of the 2005 Commission Decision in the German *Bundesliga* case:[23]

> The decision benefits both football fans and the game. Fans benefit from new products and greater choice. Leagues and clubs benefit from the increased coverage of their games. Readily available premium content such as top football boosts innovation and growth in the media and information technology sectors. Moreover, open markets and access to content are an essential safeguard against media concentration.[24]

In the EU White Paper on Sport, published on 11 July, 2007, Article 4.8 provides on the subject of the collective selling of sports television rights as follows:

> The application of the competition provisions of the EC Treaty to the selling of media rights of sport events takes into account a number of specific characteristics in this area. Sport media rights are sometimes sold collectively by a sport association on behalf of individual clubs (as opposed to clubs marketing the rights individually). While joint selling of media rights raises competition concerns, the Commission has accepted it under certain conditions. Collective selling can be important for the redistribution of income and can thus be a tool for achieving greater solidarity within sports.

And pertinently adds:

> The Commission recognises the importance of an equitable redistribution of income between clubs, including the smallest ones, and between professional and amateur sport.

The essential point here is that the EU will be looking for a "robust solidarity mechanism" to be put in place by sports bodies in order for them to justify the selling of their television rights on a collective basis! In other words, paying "lip service" to the idea/concept of "sporting solidarity" will not be sufficient for Competition Law clearances.

4 The FIFA and UEFA "Crown Jewels of Sport" Television Cases

For the sake of completeness, it should be mentioned that the sale of sports television rights in the European Union is not only subject to restraints imposed by the EU Competition Rules, for the benefit of consumers of sports programmes, but also to the important provisions of the so-called "Television Without Frontiers" EU Directive.

On 17 February 2011, the General Court (formerly the Court of First Instance) (Seventh Chamber) of the European Court of Justice (ECJ) handed down two landmark Judgements in the so-called "Crown Jewels" cases brought by FIFA, the World Governing Body of Football, and UEFA, the European Governing Body of Football.

23 COMP/C.2/37.214.
24 IP/05/62, 2005-01-19.

Article 3a of Council Directive on the coordination of certain provisions laid down by law, regulation or administrative action in member states concerning the pursuit of television broadcasting activities,[25] generally known in shorthand as the "Television Without Frontiers" Directive, provides in paragraph 1 as follows:

> (1) Each Member State may take measures in accordance with Community law to ensure that broadcasters under its jurisdiction do not broadcast on an exclusive basis events which are regarded by that Member State as being of major importance for society in such a way as to deprive a substantial proportion of the public in that Member State of the possibility of following such events via live coverage or deferred coverage on free television. If it does so, the Member State concerned shall draw up a list of designated events, national or non-national, which it considers to be of major importance for society. It shall do so in a clear and transparent manner in due and effective time. In so doing the Member State concerned shall also determine whether these events should be available via whole or partial live coverage, or where necessary or appropriate for objective reasons in the public interest, whole or partial deferred coverage.

In the present cases, The United Kingdom and the Belgian Governments had decided to list all the FIFA World Cup matches and the UEFA European Championship as sporting events considered to be of "major importance for society" in the UK and Belgium and, as such, should, therefore, be shown on "free-to-air" television. The effect of these listings was to preclude these sports events from being sold exclusively to subscription and "pay-per-view" channels, thus affecting the revenues of the sports bodies concerned.

FIFA and UEFA argued that the listing of these events, which are money-spinners for them, as "free-to-air" under the "Television Without Frontiers Directive" restricted their bargaining rights with television companies for football content, which, as noted in this article, is much and widely in demand, and were contrary to EU Law; in particular, the Competition Rules under the renumbered articles 101 and 102 of the Lisbon Treaty.

The ECJ held that the World Cup and the European Championship were single sporting events and could not, therefore, be divided up (known, in the jargon, as "siphoning off") at the will of FIFA and UEFA. In other words, a "pick and mix" approach could not be followed. The Court also held that the governmental measures, taken after full public consultation, to list these events as ones to be broadcast on "free-to-air" television were proportionate and served the public interest; and, moreover, did not go any further than was necessary to attain that objective. In other words, they were not anti-competitive and, therefore, compatible with EU Law.

---

25 Council Directive 89/552/EEC of 1989-10-03 (OJ 1989 L 298 23), as inserted by Directive 97/36/EC of the European Parliament and of the Council of 1997-06-30 amending Directive 89/552/EEC (OJ 1997 L 202 60).

Not surprisingly, FIFA and UEFA have appealed against this ruling to the Grand Chamber of the ECJ. On 12 December, 2012, Niilo Jaaskinen, the Advocate General in these cases, rendered an Opinion to dismiss the FIFA and UEFA Appeals, supporting the Rulings of the ECJ.[26]

The Court normally follows the Opinion of the Advocate General, when rendering its final judgement, which the Court did in the FIFA and UEFA Appeal Cases, dismissing them on 18 July, 2013 "in their entirety".[27] In particular, the Court held that it was "for the member states alone to determine the events which are of major importance" to their viewing public; and also that the tournaments (the World Cup and the Euros) "in their entirety have always been very popular among the general public and not only viewers who generally follow football matches on television."

Thus, the listing by the UK and Belgian Governments of the entire FIFA World Cup and the entire UEFA 'EUROS' as "free-to-air" sporting events under the "Television Without Frontiers" Directive was upheld by the European Court.

5 Concluding Remarks

This exciting and technology-led brave new media world[28] is undoubtedly having an appreciable effect on the sporting world in general and, in particular, presents further challenges to sports broadcasting rights holders, sports governing bodies, sports persons and teams and other stakeholders, as well as their professional advisors, not least their lawyers.

This inevitably leads to all kinds of conflicts that will need to be resolved by public authorities – at the national and supra-national levels – and sports bodies and administrators themselves. As Weatherill has pertinently observed:

> [s]ome of the most intriguing tensions in the years to come are likely to centre on the attempts of governing bodies to satisfy the commercial aspirations of the most powerful participants while also maintaining vertical

26 See CJEU Press Release No 164/12.

27 For the full texts of the CJEU judgements in these Appeal Cases, see C-201/11 P, C-204/11P & C-205/11P, which are published on the official website of the Court at 'www.curia.europa.eu'.

28 See Report by Market Analysts Forrester entitled, "The Battle For The Digital Home", referred to in Fry "Delivering Outside The Box" June 2005 *SportBusiness International* 41. In this Report, various scenarios that might arise in the so-called "converged sports world" are suggested, including a possible "joint venture between the NFL and Fox to create an immersive experience where sports fans can choose viewer-selected camera angles and split-screen content ... and Disney's acquisition of Electronic Arts. With EA's sports games, Disney/ESPN becomes the premier sports brand on PCs, TV and consoles".

> solidarity within the sport and preserving the broader integrity of the character of the event.[29]

This is a rather difficult balancing act to perform. As Weatherill further points out:

> [t]he prominence of EC law's intervention in sport in recent years is above all the consequence of the 'commercialisation' of the sector, in particular as a result of its close association with the helter-skelter development of the broadcasting industry. In fact, much of the economically significant sports-related material that tumbled into the Commission's in-tray in the late 1990s was concerned directly or indirectly with broadcasting. In some respects the Commission's recent preoccupation with sport has been driven by its need to monitor the commercially much more important broadcasting sector, in which it is profoundly anxious to forestall practices that will facilitate existing incumbents' anxiety to impede new entrants. And it is highly plausible that the pace of technological change will increasingly throw up new forms of rapid mass communication, generating intensified fragmentation in the pattern of audiovisual services. This will fuel yet more demand for rights to broadcast sports events, and bring with it yet more challenges for EC competition law.[30]

Therefore, interesting and challenging times lie ahead for all those involved in any way and at any level in the sports broadcasting field, not least concerning the world's favourite game, football. In this respect, it will be interesting to see how the law at the EU and the national levels develops, including country initiatives, like the new Sports Broadcasting Regulations in Italy, designed to ensure transparency and efficiency of the broadcasting rights market, while improving the competitive balance among participating clubs. A continuing dichotomy and conflict in the sports law field.

The sale of television and new media rights in respect of national and international sports events already provide a significant source of revenue to international and national sports federations alike, and will, no doubt, continue to do so for the foreseeable future, especially for commercial exploitation on new media platforms, notwithstanding the present economic climate in the global economy. Indeed, hardly a day passes without some announcement of a major sale of television rights having been made in respect of a particular sporting event somewhere in the world.

Despite all this economic preoccupation and doom and gloom in the financial markets, sport is now a multi-billion dollar industry worldwide and will continue to be so. As such, some would argue, that perhaps sport has lost touch with its Corinthian roots and values, as well as its Olympian ideals and *raison d'etre*, since nowadays it seems to be the winning rather than the taking part that motivates sports persons and teams and what really counts – and there is certainly a lot to play for!

29 Weatherill *European Sports Law Collected Papers* (2007) 246.
30 *Idem* 296.

Nevertheless, it must be said that, without the mega sums generated by the sale of television rights, sporting spectaculars like the Olympics, the FIFA World Cup and the UEFA Champions' League (to mention but a few) could not take place – the costs of organising and staging them would be prohibitive. And, accordingly, athletes and sports fans alike throughout the world would be the losers!

Diverse probleme rondom die bestaan en geldigheid van 'n testament by die dood van die testateur (deel 1)

Linda Schoeman-Malan
BA LLB LLD
Professor Departement Privaatreg, Universiteit van Pretoria

SUMMARY

Miscellaneous Problems Relating to the Existence and Validity of a Will at the Death of the Testator

Executing a will is an important step in estate planning. A will gives the testator the opportunity to bequeath his assets and to indicate how he wants his estate to be distributed after his death. Not only is the proper execution of a will important but it should be updated regularly and carefully to provide for changed circumstances. The will must accordingly be readily available for the testator to amend or revoke. The availability of the will after the testator's death is crucial in order for the administration process to commence. Although an easily accessible will can lead to problems surrounding the authenticity of the will, it can contribute to the prompt administration of the estate as it must be submitted to the master in terms of the Administration of Estates Act.

In this article numerous problems relating to the existence, availability and recovery of a will at the death of the testator are discussed. Interrelated issues such as forgery, lost wills, missing wills, concealment of wills and revocation by destruction of a will are discussed with reference to case law and other examples. Questions as to the proper custody of the will, before and after the death of the testator, resolving disputes amongst next-of-kin, and the onus of proof, are discussed. Cases, that in fact dealt with common law lost wills, but where section 2(3) condonation applications were incorrectly brought, are discussed. Possible resolutions for the miscellaneous problems with wills are explored and recommendations made to solve potential problems.

1 Inleiding

In die erfreg is geskille tussen naasbestaandes oor die testament en erfporsies nie onbekend nie en lei dikwels tot uitgerekte litigasie tussen belanghebbendes.[1] Die feit dat formaliteitsvereistes vir verlyding van testamente redelik eenvoudig, en die inhoud van 'n testament vertroulik is en geheim gehou kan word,[2] skep geleentheid vir bedrog tydens die

1 Corbett *The Law of Succession in South Africa* (2001) 113; Abrie *Bestorwe Boedels* (2011) 87.

2 Die erflater kan selfs sy testament in sy eie handskrif skryf. Sien Kahn *Bloody Hand: Wills and Crimes* (2003) "Forgery and falsification of a will" 152 ev. In *Raubenheimer v Raubenheimer* 2012 5 SA 290 (HHA) par 1 word die volgende opmerking gemaak: "*It is a never-ending source of amazement that so many people rely on untrained advisors when preparing their wills, one*

verlyding, bewaring en herroeping van testamente.[3] Die artikel handel met moontlike knelpunte rakende die testament, wat hulle mag voordoen tussen potensiële erfgename by die dood van 'n erflater.[4] Die belang daarvan om 'n geldige testament na te laat,[5] die testament te bewaar (voor en na die dood van die erflater),[6] asook moontlike scenario's wat by die dood kan voorkom, soos bewerings van bedrog, vervalsing, verberging of vermissing van 'n testament word ook onder die loep geneem.[7] Die aanwending van artikel 2(3) van die Wet op Testamente[8] wat handel met kondonasie van nie-nakoming van formaliteitsvereistes word kortliks bespreek ten einde onduidelikhede op te klaar oor die aanwending daarvan by vermiste testamente.[9] Die omvang van probleme en die bewyslas word aan die hand van regspraak geïdentifiseer. Ter toeligting word ook na ongerapporteerde sake, (wat in dagblaaie bespreek is)[10] voorbeelde en ander regstelsels verwys. Gevalle waar die testament verlore raak na die dood (tydens die beredderingsproses) word ook in oënskou geneem. Ten slotte word moontlike oplossings met betrekking tot vraagstukke rondom die betwiste testament aan die hand gedoen.[11]

2 Die Belang Daarvan om 'n Geldige Testament na te laat

2 1 Algemeen

Die verlyding van 'n testament is 'n belangrike en noodsaaklike stap wat 'n persoon tydens sy lewe neem as 'n integrale deel van

of the most important documents they are ever likely to sign."

3 Vgl Kahn "The will that won't – Formalities gone bananas" in *Huldigingsbundel Paul van Warmelo* (ed Van der Westhuizen) (1984) 128-130; Corbett ea 113.

4 Vgl oa *Van Wetten v Bosch* 2004 1 SA 348 (HHA); *Blom v Brown* 2011 3 All SA 223 (HHA); *Pienaar v Master of the Free State High Court, Bloemfontein* 2011 6 SA 338 (HHA); *Reichman v Reichman* [2011] ZAGPJHP; *Raubenheimer v Raubenheimer* 2012 5 SA 290 (HHA).

5 Par 2.

6 Par 3.

7 Par 5-7.

8 7 van 1953.

9 Par 9.

10 Vgl oa Van Rooyen "Jare lange twis oor erfporsie tot 'n einde" *Volksblad* (2013-02-17); Jansen "Die 'madam' en die R110 Miljoen" *Rapport* (2012-02-25); Krige "Atlantis-vrou in hof oor bedrog en moord" *Son* (2007-01-23); Carstens "Sakeman vervals glo vrou se testament sodat hy kan erf" *Beeld* (2002-05-16); Tau "Mystery Surrounds Lolly's Will" beskikbaar by www.iol.co.za/.../mystery-surrounds-lolly-s-missing-will-1 (besoek 2012-07-22).

11 In die bespreking wat volg word veronderstel dat die tersaaklike testament ingevolge a 2(1)(a) van die Wet op Testamente 7 van 1953 aan die formaliteitsvereistes voldoen of voldoen het. Sien ook Van der Merwe & Rowland *Die Suid-Afrikaanse Erfreg* (1990) 149; Corbett ea 49; De Waal & Schoeman-Malan *Erfreg* (2008) 54.

boedelbeplanning.[12] Die belangrikheid daarvan om 'n testament op te stel, lê daarin dat dit aan 'n erflater die geleentheid bied om na sy dood oor sy bates te beskik en aan te dui hoe hy wil hê sy boedel hanteer moet word.[13] 'n Persoon kan self sy testament opstel, alhoewel aanbeveel word dat van die dienste van 'n professionele testamentopsteller gebruik word.[14] Ingevolge die beginsel van testeervryheid kan 'n persoon sy bates nalaat aan wie hy wil.[15] Geen persoon (naasbestaande of familielid) kan daarop aanspraak maak om testamentêr bevoordeel te word nie.[16] Die behoorlike verlyding van 'n testament is belangrik maar so ook die opdatering en bewaring daarvan.[17] Die testateur het verder ook die bevoegdheid om sy testament voor sy dood te wysig of te herroep.[18] By die dood van die erflater moet sy boedel sonder versuim by die meester ingehandig word ingevolge artikel 8 van die Boedelwet.[19]

Daar sal dan bepaal word of die erflater testaat (met 'n geldige testament) of intestaat (sonder 'n geldige testament) gesterf het.[20] Indien 'n geldige testament nagelaat word, moet daaraan uitvoering gegee word. Indien die egtheid of geldigheid van "dié testament" egter betwyfel of betwis word sal die hof bepaal welke testament die laaste geldige

12 De Waal & Schoeman-Malan 82 & 93; Abrie ea *Bestorwe Boedels* (2011) 54; Pace & Van der Westhuizen *Wills and Trusts* (2012) A1-2; Nathan "Careful planning of your last will & testament today secures your wishes for tomorrow" 2012 *e-bizzline* beskikbaar by http://www.gt.co.za/publications/ (besoek 2013-03-05); *Raubenheimer v Raubenheimer* 2012 5 SA 290 (HHA).

13 Wiechers *Testamente 'n kortbegrip* (1988) 5; Sien ook De Waal "A Comparative Overview" in *Exploring the Law of Succession* (red Reid ea) (2007) 6-9; Abrie ea 54 & 56; De Waal & Schoeman-Malan 4; Pace & Van der Westhuizen A 42; Corbett ea 34; Jooste "Where there's a will, there's a way" 2010 *Finweek* 48.

14 Sien *Raubenheimer v Raubenheimer* 2012 5 SA 290 (HHA) par 1; Kahn (1984) 130; Wiechers 12-13 & 17; Senekal "Met registrasie raak testament nie weg" *Rapport* (2002-09-15) 11; Nathan 2012 *e-bizzline.*

15 Corbett ea 39. Sien ook Jansen "Die 'madam' en die R110 Miljoen" *Rapport* (2012-02-25), waar die erflater sy sewe kinders onterf en sy boedel verdeel het tussen sy boekhoudster, prokureur en broerskind; Prince "Die Volle Waarheid oor Petersen se Boedel" *Die Burger* (2007-05-18) waar die skrywer rapporteer oor 'n bekende kunstenaar wat in 1985 'n testament gemaak het waarvan sy eerste vrou, kinders en tweede vrou nie geweet het nie. Verder Tau "Mystery Surrounds Lolly's Will" beskikbaar by www.iol.co.za/.../mystery-surrounds-lolly-s-missing-will-1 (besoek 2012-07-22) waar aangevoer word dat wyle Lolly Jackson en sy vrou vervreemd was en dat hy vir 'n sakevennoot voorsiening gemaak het in 'n latere testament wat nie gevind kan word nie.

16 Corbett ea 7; Abrie ea 74.

17 Nathan 2012 *e-bizzline;* King "Pitfalls To Avoid" *Personal Finance* 2012 beskikbaar by http://www.iol.co.za/business/personal–finance/financial–planning/estate/wills–pitfalls–to–avoid–1.1343476 (besoek 2012-08-08) wys daarop dat dit nodig is om jou testament meer gereeld te hersien en nie net wanneer lewensveranderende omstandighede plaasvind nie.

18 Vir wysiging en herroeping van 'n testament sien Corbett ea 69 & 88; De Waal & Schoeman-Malan 82 & 93.

19 66 van 1965. Vgl ook Abrie ea 110 ev; Sonnekus "Freedom of Testation and the Ageing Testator" in Reid ea (2007) 92; Corbett ea 115.

20 *S v Van Zyl* 1985 3 SA 25 (A); *S v TS Maqubela* (onafgehandelde-saak WKH); *Uys v Uys* [2008] ZANCHC 30; *Marais v Botha* [2008] ZAWCHC 111.

testament is.[21] Indien aangevoer word dat daar wel 'n testament verly is wat by die dood van die testateur nie gevind kan word nie, ontstaan die vraag na die status van die testament, naamlik of dit verberg, herroep, of verlê, per abuis vernietig, of in die bewaring van 'n onbekende persoon of bloot op 'n onbekende plek gelaat is.[22]

Die onderliggende belang van die testate erfreg en die verlyding van 'n testament is dat daarna gestreef word om aan die ware bedoeling van die testateur, soos verwoord in sy testament, as sy laaste wilsbeskikking uiting te gee.[23] Indien die testament behoorlik verly en geredelik beskikbaar is, vereenvoudig dit die beredderingsproses.[24] Indien die testament betwis word of nie gevind kan word nie, kan die beredderingsproses onnodig vertraag word.[25]

2 2 Die Testament en die Beredderingsproses

Die bestaan al dan nie van 'n geldige testament speel 'n deurslaggewende rol in die afhandeling van die boedel.[26] Dit is verder belangrik dat die testament so gou moontlik na die dood van die testateur opgespoor word,[27] aangesien daar sekere statutêre bepalings is waaraan

21 *Haribans v Haribans* [2011] ZAKZPHC 46 par 41; *Scott v Master of the High Court, Bloemfontein* [2012] ZAFSHC 190 par 16; Corbett ea 117.

22 Die "regstatus van die testament" beteken in hierdie konteks die bestaan, verlyding, geldigheid, herroeping ens van die testament, derhalwe welke testament is die laaste *geldige* testament.

23 De Waal & Schoeman-Malan 231; Sonnekus (2007) 92. Wiechers 10 verwys na die verlyding van die testament as 'n kommunikasieproses. Vgl Cloete "Guard against disappointed beneficiaries: the will drafting duties of legal practitioners" 2003 *TSAR* 540 ev.

24 Sien Cronjé & Roos *Casebook on the Law of Succession/Erfregvonnisbundel* (2002) 87; Probleme met betwiste testamente kom algemeen wêreldwyd voor: Biggs "Losing It" 2005 *The Step Journal* 28; Duhaime "The Lost Will" 2007 *Legal Resources* beskikbaar by http://www.org/LegalResources/ElderLawWillsTrustsEstates/LawArticle–270/The–Lost–Will.aspx. (besoek 2012-09-22). Vgl verder Jacobs "The Lost or Missing Testamentary Instrument" *The Electronic J of the Bar Ass Queensland* beskikbaar by hhtp://www.hearsay.org.au (besoek 2012-08-14); Jooste 2010 *Finweek* 48.

25 Vir vb van uitgerekte litigasie vgl *Van Wetten v Bosch* 2004 1 SA 348 (HHA); *Blom v Brown* 2011 3 All SA 223 (HHA); *Pienaar v Master of the Free State High Court, Bloemfontein 2011 6 SA 338 (HHA); Raubenheimer v Raubenheimer* 2012 5 SA 290 (HHA). Sien ook die Suid-Australiese sake *In the Estate of Hall (deceased)* 2011 SASC 117 en *Thorner v Major* 2009 UKHL 18.

26 Nathan 2012 *e-bizzline*; Cloete 2003 *TSAR* 544; Horace "Beplan jou Testament" *Bloemfontein courant* (2012-08-01) skryf: "'n Gebrekkige testament is by magte om jare se harde werk en finansiële beplanning met jou afsterwe tot niet te maak. 'n Testament kan egter ook die bewaarder wees wat jou rykdom beskerm, en regverdig verdeel tussen jou erfgename." Beskikbaar by http://www.bloemfonteincourant.co.za/article/128/Beplan-jou-Testament (besoek 2013-04-15).

27 Senekal 2002 *Rapport* 11. Vgl ook Certainty "National Wills register and Will Services" beskikbaar by http://www.certainty.co.uk/ (besoek 2012-09-18). Cloete 2003 *TSAR* 544 wys daarop dat die testamentopsteller twee take het: Die een is om die testament veilig te bewaar, en die ander om die bestaan daarvan te openbaar.

voldoen moet word.[28] Die beredderingsproses neem spoedig na die dood van 'n persoon 'n aanvang en die naasbestaandes of eksekuteur moet die dood aanmeld.[29] Die laaste geldige oorspronklike testament moet ingehandig word ten einde te bepaal wie die bevoordeeldes is. Die proses word deur die Boedelwet[30] voorgeskryf. Artikel 8 van dié Wet handel met die aanstuur of aflewering van 'n testament en die aanvaarding en registrasie daarvan deur die meester.[31] Ingevolge artikel 8 is daar 'n verpligting op naasbestaandes om enige "testamentêre geskrif"[32] by die meester in te handig. Artikel 8 bepaal soos volg:

> (1) Enigiemand wat 'n dokument wat 'n testament is of heet te wees, in sy besit het ten tyde van of te eniger tyd na die dood van enige persoon wat die dokument verly het, moet, sodra hy van die dood te wete kom, die dokument aan die Meester stuur of aflewer.[33]

Die persoon wat die testament vind, of wat die testament bewaar, kan dus nie self oor die geldigheid daarvan besluit nie.[34] Indien daar 'n geskil oor die bestaan of geldigheid van die dokument wat ingehandig is, ontstaan, of indien daar meerdere testamente ingehandig word, sal die beredderingsproses noodwendig onderbreek word.[35] Indien dit vir die meester blyk dat so 'n dokument (wat as testament ingehandig is) om die

---

28 Die beredderingsproses is dieselfde vir die testate en intestate erfreg.

29 Abrie ea 110; Pace & Van der Westhuizen A12. Die boedel moet binne 14 dae aangemeld word. Sien ook FISA "Original Will Lost" beskikbaar by http://fidsa.org.za/original–will–lost/2010 (besoek 2010-03-04). Sien ook *Haribans v Haribans [2011] ZAKZPHC 46* [2011] ZAKZPHC 46 par 11 waar die hele proses van aanmelding en registrasie van testamente verduidelik word. In *Theart v Scheibert* [2012] ZASCA 131 word die dood eers sewe jaar later aangemeld: "*The testatrix died on 11 February 1990. Two death notices were filed with the Master in terms of s 7 of the Administration of Estates Act 66 of 1955 – one by the testator, and one by Ms M M Brink who described herself in the notice she filed as a nurse/friend and who the appellant asserts was the testator's then girlfriend. The death notices are each dated 4 December 1997, ie more than seven years after the testatrix's death. Both notices stated – incorrectly – that the testatrix was married by ante-nuptial contract and that she had died intestate.*"

30 66 van 1965. Vgl reg 5.

31 Aanvaarding en registrasie van die testament na die dood van die erflater is 'n administratiewe taak en bepaal nie die geldigheid van die testament nie. Vgl *Giles v Henriques* 2008 4 SA 558 (K); *Haribans v Haribans* [2011] ZAKZPHC 46 n 21 *supra*; *Reichman v Reichman* [2011] ZAGPJHP]; *Theart v Scheibert* [2012] ZASCA 131 par 1 en 8; Kahn *Supplement to the Law of Succession in South Africa* (1994) 31; Sonnekus (2007) 92 en Hewson "Registering Your Last Will and Testament: What's the Point?" 2011 beskikbaar by http://legalwills.wordpress.com /author/timhewson/ (besoek 2013-03-06).

32 A 1 van die Wet op Testamente omskryf 'n testament as enige kodisil of testamentêre geskrif.

33 Corbett ea 114-115; Abrie ea 113. Die meester kan weier om die testament te aanvaar indien dit nie behoorlik verly, gewysig of herroep is nie.

34 Cloete 2003 *TSAR* 544. Vgl ook Rogers "The action of the disappointed beneficiary" 1986 *SALJ* 603 ev.

35 Sien n 4 en 21 *supra*; *Mabase v Dlamini* [2007] ZAGPHC 199; *Scott v Master of the High Court, Bloemfontein [2012] ZAFSHC 190 of the High Court, Bloemfontein* [2012] ZAFSHC 190; *Smith v Sampson* [2013] ZAWCHC 11.

een of ander rede ongeldig is, kan die meester ondanks die registrasie van die dokument weier om dit vir doeleindes van die bereddering van dié boedel te aanvaar alvorens die hof oor die geldigheid daarvan beslis het.[36] Die hooggeregshof sal genader moet word met 'n aansoek om die testament as die finale en geldige testament te verklaar.[37] Indien die testament op welke gronde ook al betwis word behoort die meester ingelig te word van die voorneme om 'n aansoek te bring. Die meester word gewoonlik as 'n party tot die geding gevoeg.[38] Die beredderingsproses sal dan outomaties opgeskort word hangende die uitspraak van die hof.[39]

Probleme met testamente by die dood van 'n persoon is nie onbekend nie en dit is klaarblyklik aan die toeneem.[40] Alhoewel sake nie dikwels in die hofverslae gerapporteer word nie,[41] verskyn daar gereeld

36 A 8(4) Boedelwet.

37 Sien *Smith v Sampson [2013] ZAWCHC 11* [2013] ZAWCHC 11 par 12. Vgl verder *Lipchick v Master of the High Court* [2011] ZAGPJHC 49*; Marais v Botha* [2008] ZAWCHC 111; *Levin v Levin* [2011] ZASCA 114*; Ex parte Porter* 2010 5 SA 546 (WK); *Ex parte Warren* 1955 4 All SA 352 (W). Sien ook King *Personal Finance* 2012-aanlyn; Senekal 2002 *Rapport* 11; Fourie "Kan jou testament betwis word?" *PSGKonsult* beskikbaar by http://www.psgkonsult.co.za/ 2011 (besoek 2012-09-12).

38 *Reichman v Reichman* [2011] ZAGPJHP par 12: "*The first respondent cannot be a judge in his own cause and cannot rely on the Master to resolve this factual dispute. If the parties are unable to resolve the dispute among themselves only a court of law would be able to do so*". Vgl ook *Theart v Scheibert* [2012] ZASCA 131 par 1 en 8; *Pienaar v Master of the Free State High Court, Bloemfontein* 2011 6 SA 338 (HHA); *Lipchick v Master of the High Court* [2011] ZAGPJHC 49 par 1*; Taylor v Taylor* 2012 3 SA 219 (OK); *Yokwana v Yokwana* [2013] ZAWCHC 22 par 3.

39 Abrie ea 98; *Ex parte Porter* 2010 5 SA 546 (WK); *Yokwana v Yokwana* [2013] ZAWCHC 22 par 7; *Blom v Brown* 2011 3 All SA 223 (HHA)*; Lipchick v Master of the High Court* [2011] ZAGPJHC 49*; Pienaar v Master of the Free State High Court, Bloemfontein* 2011 6 SA 338 (HHA)*; Levin v Levin* [2011] ZASCA 114 par 9*; Raubenheimer v Raubenheimer* 2012 5 SA 290 (HHA).

40 Vgl oa *Ex parte Porter* 2010 5 SA 546 (WK); *Van Wetten v Bosch* 2004 1 SA 348 (HHA); *Uys v Uys* [2008] ZANCHC 30; *Marais v Botha* [2008] ZAWCHC 111; *Lipchick v The Master* [2011] ZAGPJHC 49; *Haribans v Haribans* [2011] ZAKZPHC 46*; Mabase v Dlamini* [2007] ZAGPHC 199; *Scott v Master of the High Court, Bloemfontein* [2012] ZAFSHC 190; *Smith v Sampson* [2013] ZAWCHC 11; *Reichman v Reichman* [2011] ZAGPJHP; *Theart v Scheibert* [2012] ZASCA 131; Cronjé & Roos 47-51; De Waal & Schoeman-Malan 125-126; Gausden & King *Wills: a Practical Guide* (2011) 1. Sien ook die Suid-Australiese sake *In the Estate of Hall (deceased)* 2011 SASC 117 en *Thorner v Major* 2009 UKHL 18. Sien ook Masondo & Basson "Madiba: Dogters mik na R25m." *Rapport* (2013-04-13) waar die skrywers opmerk: "Dit is nie ongewoon vir families om oor erfporsies te stry nie. Maar om te twis oor die erflating van iemand wat nog leef? Ongehoord!"

41 Sien die ongerapporteerde saak *Christina van der Watt v Die Meester van die Hooggeregshof Pretoria* Saaknr 39934/05 (2007-02-15); Sien verder die twee ongerapporteerde sake van vervalsing in die 1990's teen Eric Stafford *Natal Sunday Star* (1993-10-22) 1 en (1993-10-24) 5. Vlg ook Du Plooy-Gildenhuys & Blackbeard "The Sabadia Conspiracy" 1999 *THRHR* 148 waar die saak *S v Manyape* (ongerapporteerde saak 159/97 TPD) bespreek word.

nuusberigte in dagblaaie in die verband.[42] Artikel 102 van die Boedelwet bevat sekere strafbepalings vir gevalle waar daar by die dood van 'n testateur vuilspel in verband met die verlyding van die testament of die testament wat nie opgespoor kan word nie, vermoed word. Ingevolge artikel 102 is iemand wat 'n dokument wat 'n testament heet te wees, steel, opsetlik vernietig, verberg, vervals of beskadig aan 'n misdryf skuldig, en by skuldigbevinding aan so 'n oortreding kan 'n boete of tronkstraf opgelê word.[43] Ten spyte van laasgenoemde artikel word in die praktyk klaarblyklik eerder gesteun op 'n klag van bedrog as op 'n oortreding ingevolge artikel 102.[44]

3 Die Status van die Testament

3 1 Is Daar 'n Testament Verly?

Die verlyding van 'n testament is 'n belangrike aangeleentheid en moet met erns bejeën word. Op die webblad *Ehow*[45] word soos volg opgemerk:

> *Dealing with the loss of a loved one can be difficult. To obtain a copy of a last will and testament, one must be sure a last will and testament exists or has not been lost. A court of law considers a last will and testament as one of the most sacrosanct documents, because it is the only way for the departed to communicate or express his wishes for the remains of his estate. A will is also considered the most private of official papers. It is not available for public record until almost a year after the departed has passed away, with strict laws regarding how the will is carried out. Any potential heir has the right to obtain a copy of a last will and testament before it becomes public record.*

Na verlyding van 'n testament is dit eerstens belangrik dat die testament *tydens die lewe van die testateur*, geredelik toeganklik is, sodat dit gewysig kan word om by veranderde omstandighede aan te pas.[46] Tweedens is

42 Vgl n 15 en Van Rooyen "Jare lange twis oor erfporsie tot 'n einde" *Volksblad* (2013-02-17); Krige "Atlantis-vrou in hof oor bedrog en moord" en "Carmen Fortuin in die hof oor bedrog en moord op man" *Son* (2007-01-23); Jansen "Die 'madam' en die R110 Miljoen" *Rapport* (2012-02-25); Carstens "Sakeman vervals glo vrou se testament sodat hy kan erf" *Beeld* (2002-05-16) 6; Botha "Testament glo 'gekook'" *Volksblad* (2012-04-15) waar polisiebeamptes aangekla word van vervalsing deurdat hulle die oorledene se duimafdruk verkry het nadat hy 'n beroerte gehad het. Sien verder Kahn (2003) "Forgery and falsification of a will" 152 ev.

43 'n Boete of gevangenisstraf vir 'n tydperk van hoogstens sewe (7) jaar kan opgelê word. In *S v Van Zyl* 1985 3 SA 25 (A) is die beskuldigde skuldig bevind aan 'n oortreding van a 102 en tot drie jaar gevangenisstraf gevonnis waarvan die helfte opgeskort is. Vgl ook Sonnekus "Voorgestelde statutêre wysigings van die Erfreg" 1992 *TSAR* 159.

44 Sien par 5 1 en *S v Van Zyl* 1985 3 SA 25 (A); Kahn (2003) 152 ev.

45 Eie kursivering. Beskikbaar by http://www.ehow.com/how_4779918_obtain-copy-last-testament.html (besoek 2013-04-02).

46 Pace & Van der Westhuizen A 42 merk op: "*But a testator should not relax once he has executed his will. It should also be revised as often as changes are necessitated by a change in his personal circumstances, for example, marriage,*

dit belangrik dat die testament *na die dood van die testateur* beskikbaar is, sodat die beredderingsproses glad kan verloop. Onsekerheid by die dood oor die vraag of die oorledene 'n testament nagelaat het en waar dit gehou word, kan onnodige komplikasies[47] meebring. Senekal[48] waarsku: "Baie mense se planne, polisse en testament word nooit gevind nie. Pleks daarvan bly dit 'weggesteek' by verskillende maatskappye in lêers, skoendose of ander geheime plekke."

Natuurlik wil potensiële erfgename nie die indruk skep dat die testateur se dood afgewag word nie, maar dit is tog in belang van alle belanghebbendes[49] en vriende dat sekere basiese vrae gestel word ten einde sekerheid te kry oor die vrae of die oorledene 'n testament het en waar dit gehou of bewaar word.[50] Dit staan die testateur vry om nie die inhoud van sy testament te openbaar nie.[51] Selfs die getuies hoef nie te weet wat in die testament staan nie.[52] Die bekendmaking van aspekte rondom 'n testament kan egter soos 'n tweesnydende swaard wees. Aan die een kant moet die erflater seker maak dat sy testament toeganklik is ten einde te verseker dat die testament na sy dood gevind word maar aan die ander kant is dit so dat indien dit geredelik toeganklik is, die testament maklik vernietig of verberg kan word deur diegene wat nie tevrede is met die bepalings van die testament nie.[53]

---

birth of a child, replacement of assets, a beneficiary predeceasing the testator or a divorce, to mention but a few or by legislation or case law." Sien *Marais v The Master* 1984 4 SA 288 (D) waar die oorspronklike van die testament wat die testeur wou herroep in besit was van die prokureur terwyl die testateur net 'n afskrif gehad het wat hy wou herroep. Vgl verder Corbett ea 114.

47 Vgl *Ex parte estate Davies* 1957 3 SA 471 (N) 472-473 waar die erflater die identiteit van 'n begunstigde in 'n geseëlde koevert geopenbaar het en *Levin v Levin* [2011] ZASCA 114 waar die oorledene 19 verskillende testament verly het. Vgl verder Potgieter "Hy onterf hulle: nou is die as weg" *Beeld* (2012-02-24); King *Personal Finance* 2012-aanlyn.

48 2002 *Rapport* 11. Vgl ook King *Personal Finance* 2012-aanlyn; Hewson 2011-aanlyn.

49 Corbett ea 117. Wat insluit die eksekuteur, naasbestaandes en ander bevoordeeldes.

50 Gunnarsson "Should you store your client's will?" 2006 *Illinois Bar J* 532 merk soos volg op: "*I know this is not always easily done without appearing 'greedy', but it will be extremely difficult to carry out the wishes of the person who made the will if that person does not let someone know where the will is.*" Sien ook Burger "*How to find a lost will*" 1999 beskikbaar by http://www.burger.///willfind.htm (besoek 2012-09-02); Verder ook Shakespeare "*Lost Wills*" beskikbaar by http://www.access-legal.co.uk/legal-news/lost-wills-lu-2172.htm 2010 (besoek 2012-10-13) en Kahn (2003) viii.

51 Vgl n 15 *supra*.

52 Die getuies bevestig slegs die handtekening van die erflater en nie die inhoud van die testament nie. Sien *Melville v The Master* 1984 3 SA 387 (K) 395; De Waal & Schoeman-Malan 51; Van der Merwe & Rowland 186.

53 *Yassen v Yassen* 1965 1 SA 438 (N); *Pillay v Nagan* 2001 1 SA 410 (D); *Marais v Botha* [2008] ZAWCHC 111. King *Personal Finance* 2012-aanlyn sê die volgende: "When money is involved, some beneficiaries and potential beneficiaries will look for any loophole to exploit". Vlg Kahn (2003) se besprekig van: "*State v Van Zyl – justice was done"* op 159*; "The missing Hertzog will'* op 149*;* "Billionaire Howard Hughes-*was there a will?"* op 166.

3 2 Bewaring van die Testament

Die testateur en/of opsteller van 'n testament, moet na die verlyding daarvan seker maak dat die testament op 'n veilige plek bewaar word.[54] Dit is binne die testateur se diskresie om te bepaal wie die oorspronklike testament, of afskrif, gaan bewaar. Die testateur kan kies om die testament in sy besit te hou of om dit vir veilige bewaring aan iemand anders te oorhandig.[55] Normaalweg word 'n oorspronklike of duplikaat-oorspronklike testament deur die opsteller aan die testateur of sy gade of 'n familielid oorhandig.[56] 'n Naasbestaande of gade, aangewese eksekuteur, vennoot of vriend, behoort te weet dat daar 'n testament verly is en waar dit gehou word.[57] Corbett en andere[58] merk in die verband soos volg op: *"The safe custody of a will prior to the testator's death is of the utmost importance. Considerable uncertainty may arise where a testator's will cannot be found on the testator's death."* Die redes waarom die testament in veilige bewaring gelaat moet word, is voor die hand liggend. Na die dood van 'n erflater kan hy geen lig werp op enige vrae nie en die beredderingsproses kan nie 'n aanvang neem nie.[59] King[60] waarsku:

---

54 *Ex parte Porter* 2010 5 SA 546 (WK) par 8-9; Corbett ea 114; Fourie 2011 *PSGKonsult*-aanlyn; Nathan 2012 *e-bizzline*. Biggs 2005 *The Step Journal* 28 merk op: "*Avoiding the loss of the original will in the testators lifetime*" raai testamentopstellers aan om die testament nie in die sorg van die testateur te laat nie. Pace & Van der Westhuizen A20 "*During his lifetime, however, the will can be kept by the testator or by some other person or institution for example his attorney or his banker. Since a will is a most important document it must be safely preserved during the lifetime of the testator*".

55 *In re Beresford* 2 SC 303; *Lipchick v Master of the High Court* [2011] ZAGPJHC 49 waar die handgeskrewe dokument aan 'n jarelange ou vriend oorhandig is. Sien ook Cloete 2003 *TSAR* 544; Hewson 2011-aanlyn.

56 Corbett ea 114 n 6 doen aan die hand dat die testateur die testament moet bewaar en dat 'n afskrif gelaat moet word in die hande van die eksekuteur.

57 Sien ook Gray "The case of the missing will" 2012 *Step Journal* beskikbaar by http://www.stepjournal.org/journal archive (besoek 2012-09-20); Heckenberg "Lost Will" *Find law Australia* 2012 beskikbaar by http://findlaw.com.au/articles/1349/lost-will.aspx (besoek 2012-08-18).

58 113. Die skrywers verwys ook na die vermoede dat die testateur die testament vernietig het, wat in werking tree indien die testament nie gevind kan word nie. Vgl ook 99-100.

59 Biggs 2005 *The Step Journal* 28; Duhaime *Legal Resources* 2007-aanlyn. Die beskikbaarheid van die testament bespoedig die proses; Jooste 2010 *Finweek* 48; Shakespeare 2010-aanlyn. Sien ook Suid-Afrikaanse Regskommissie Projek 22 *Hersiening van die Erfreg; Wysiging en Herroeping van Testamente* (1991) par 2 158-2 163. Vgl verder *Haribans v Haribans* [2011] ZAKZPHC 46 par 1: "*The deceased, Widthit Haribans, died leaving a will dated the 21st October 2004. All parties to the application in the court* a quo *accepted that that will had been validly executed. The Master of the High Court in Durban was on the point of winding-up the estate of the deceased pursuant to this will, when a copy of a later will ('the disputed will') surfaced. It was dated the 24th June 2005. The validity of the disputed will was in issue in the court a quo*."

60 *Personal Finance* 2012-aanlyn.

> *If you get things wrong, it could have painful unintended consequences. You may simply not be able to achieve what you intended, or your beneficiaries could end up in a long and expensive court case as they battle over the interpretation of your words or actions.*

Regter Gray,[61] van die Suid-Australiese appèlhof, tydens 'n voordrag wat hy oor die onderwerp "The case of the missing will" lewer, vra die volgende vraag:

> *Do you know the location of your parents' original wills? According to research conducted by a company in the United Kingdom, two-thirds of the people surveyed did not know where to find their parents' wills.*

Onkunde by naasbestaandes oor die bestaan en bewaring van die testament sal noodwendig aanleiding gee tot 'n vertraging in die afhandeling van die boedel.

Voordat litigasie egter oorweeg word, is 'n goeie beginpunt om deeglik te soek tussen ander belangrike persoonlike dokumente van die oorledene.[62] Die testament of 'n duplikaat-oorspronklike sal gewoonlik gehou word in 'n kluis of liasseerkabinet.[63] 'n Verdere moontlikheid is dat dit na verlyding deur die opsteller ('n bank, trustmaatskappy,

61 2012 *Step Journal*. Vgl ook Duhaime *Legal Resources* 2007-aanlyn bespreking van *Re Wagenhoffer Estate* 26 SASK R.

62 *Ex parte Warren* 1955 4 All SA 352 (W); FISA 2010-aanlyn. Die Engelse saak *Cresswell v Jackson* 1860 2 F&F 24 175 ER 942 is 'n voorbeeld waar daar na bewering intensief gesoek is vir kodisille tot die testament (selfs waar niemand sou kon dink om 'n testament te vind nie) "*… but the discoveries were very strange. One codicil was in* a penny memorandum book*; the second was in a* pickle jar in a hole *in the wall under the sill of a window of an outhouse; and the third* was in that hole*"* (eie kursivering). In die saak *In the Estate of Hall* 2011 SASC 117 word bevind: "*In particular, they searched a room that contained boxes of paper and records. The deceased's parents destroyed documents that they believed to be redundant ... Further, there was a filing cabinet in the room that was said to have contained the deceased's more important documents, and some of the folders in that filing cabinet had been disturbed*". Vgl ook Leigh "How to find someone's will" *Ehow* 2012 beskikbaar by http://www.ehow.com/how 5108245 someones.html (besoek 2012-09-13); Appleby *Legal research guide to Wills and succession* (2006) 19; *Re Quinlan's Will* 1985 63 NBR 2d 429; 164 APR 429 (Probate Ct).

63 Hamill "How to find a lost will (or at least where to look)" 2009 *Estate Planning, probate/estate administration* beskikbaar by http://www.lhamillattorney.typepad.com/Massachucetts (besoek 2012-07-23). Sy verduidelik: "*I always recommend that my clients keep their original wills (and other important documents) in a fireproof box in their house, and that they let some family members know where the documents are located, and where the key is (or what the combination is) if the box is locked. This is so that when the time comes to probate the Will, which could be 20+ years from now, the original can be located*".

ouditeur, prokureurskantoor of selfs 'n individu) in bewaring geneem is.[64]

4 Problematiek by Dood van Testateur

4 1 Algemeen

Ongeërgdheid aan die kant van die testateur, deurdat hy niemand in sy vertroue neem nie, kan tot diverse probleme van uiteenlopende aard lei.[65] Uit die talle gevalle waaroor daar in hofverslae[66] en nuusberigte[67] gerapporteer word blyk dit dat onreëlmatighede rondom testamente dikwels voorkom. Kahn merk op:[68]

> *The number of relevant cases, civil and criminal, in our law reports, on falsification and forgery concerned with a will, apart from several bringing no success to the plaintiff or prosecution, of course is no reflection of the extent of the problem. Still, a few are interesting, if only as showing that there is not so queer as folks – or, I would add, so crooked, at times.*

Dit is belangrik om, indien die testament nie gevind kan word nie of onraad vermoed word, so gou moontlik regsadvies te bekom sodat die proses met betrekking tot 'n bestrede testament verduidelik kan word.[69] Testamentopstellers en prokureurs word wel dikwels geraadpleeg deur kliënte wat navraag doen oor die geldigheid van 'n bestaande testament

64 Sien n 63 *supra*. Burger 1999-aanlyn beveel aan dat ooglopende plekke soos onder die matras, in die paneelkissie of kattebak, tussen die blaaie van boeke, buite sig soos in die motorhuis of in 'n hangkas nie oorgeslaan moet word nie. Vgl ook Senekal 2002 *Rapport* 11; Leigh 2012 *Ehow*; Duhaime *Legal Resources* 2007-aanlyn; Heckenberg "Lost Will" *Find law Australia* 2012 beskikbaar by http://findlaw.com.au/articles/1349/lost-will.aspx (besoek 20120818).

65 Met onnodige regskoste tot gevolg. Vgl Pace & Van der Westhuizen A 43.

66 *Ex parte Porter* 2010 5 SA 546 (WK); *Van Wetten v Bosch* 2004 1 SA 348 (HHA); *Marais v Botha* [2008] ZAWCHC 111; *Uys v Uys* [2008] ZANCHC 30; *Lipchick v The Master* [2011] ZAGPJHC 49; *Haribans v Haribans* [2011] ZAKZPHC 46; *Mabase v Dlamini* [2007] ZAGPHC 199; *Scott v Master of the High Court, Bloemfontein* [2012] ZAFSHC 190; *Smith v Sampson* [2013] ZAWCHC 11.

67 Nuusberigte sluit in Botha "Testament glo 'gekook'" *Volksblad* (2001-04-15) waar polisiebeamptes aangekla word van vervalsing deurdat hulle die oorledene se duimafdruk verkry het nadat hy 'n beroerte gehad het; Krige "Carmen Fortuin in die hof oor bedrog en moord op man" *Son* (2007-01-23); Makwabe "Relations sour over mogul's R200m estate" 2010; Mulder "Eiendomme uit omstrede boedel van bekende entrepreneur word opgeveil" *Die Burger* (2010-11-26); Van der Westhuizen "Geen twyfel oor valse handtekening" *Beeld* (1997-09-18); Pienaar "Moord: Regter se testament vervals, beslis hof" *Beeld* (2013-01-31).

68 (1984)138.

69 Pace & Van der Westhuizen A20 "*It causes great stress amongst family members if the last will cannot be found and as has been shown above (16.1 above), if the will cannot be found following the death of the testator the will will be presumed to have been destroyed by the testator if it had been in his possession (see Corbett 113-114).*"

of bewys van 'n verlore testament.[70] Bevindinge ten opsigte van 'n betwiste testament kan 'n wesenlike invloed op die uitkoms van die verdeling van die boedel. Todd[71] waarsku praktisyns:

> *Many estate practitioners will face the situation where the original will cannot be located following the death of the testator. The destruction of a later original will can have the very real effect, if combined with dishonesty and perjury, all for the sake of money, of reviving a previous will, or leaving the estate to be distributed in accordance with the applicable intestate statute, either more favourable to the first person on the scene of death or falling upon the deceased's belongings.*

4 2 Betwiste Testamente en Litigasie

Indien 'n testament vir watter rede ook al betwis word volg daar gewoonlik lang en duur litigasie.[72] Die omvang van die probleme wat ontstaan kan uitgebreide afmetings aanneem. Gausden en King[73] merk in die verband soos volg op:

> *Litigation over disputed wills is sadly on the rise and costs can easily swallow up the whole of modest to moderate estates. Just as bad as the financial waste, are the irreconcilable rifts and misery that bitter family disputes can cause in determining capacity, claims under the Inheritance (Provision for Family and Dependants) Act 1975, mutual wills, conflicts with other death dispositions such as estoppel, foreign property and issues over the will's construction and interpretation.*

Dis moeilik om hofsake te kategoriseer omdat elke geval se feite verskil.[74] Teen die agtergrond van bogenoemde is dit nie vreemd dat Ellison Kahn die boek "Bloody Hand: Wills and Crime", waarin hy die diverse intriges en dispute wat ontstaan rondom testamente, aan die

---

70 Vgl ook "Lost wills" beskikbaar by http://chancery12.wordpress.com/2011/01/05/lost-wills/ (besoek 2012-10-10) waar gevra word: "*Does it ever happen to you that an heir shows up in your office and says something to the effect that 'Mom says you kept the original of dad's will. All we have is this (dogeared, coffee-stained, footprinted) copy,' and hands you a bedraggled handful of papyrus? Well, if it hasn't, it will*". Sien verder Biggs 2005 *The Step Journal* 28 en Heckenberg 2012 *Find Law Australia* waar die potensiële erfgename navraag doen oor vermiste testamente en "Will Forgery Claim Colapse" *Contesting Wills* (2012-02-07) beskikbaar by http://www.contesting-wills.co.uk/news-articles/will-forgery-claim-colapses.html (besoek 2013-04-11).

71 "Disinherited; Estate dispute and contested wills; Lost wills" 2010 beskikbaar by http://www.disinherited.com/article/lost–wills (besoek 2012-08-18).

72 *Reichman v Reichman [2011] ZAGPJHP* par 13: "*The final outcome of this litigation could take several years, bearing in mind that there is the possibility of an appeal. There is no certainty as to the outcome of that litigation*." Vgl ook *Yokwana v Yokwana [2013] ZAWCHC 22* par 7; *Levin v Levin* [2011] ZASCA 114 par 9; *Haribans v Haribans [2011] ZAKZPHC 46* par 41 en die sake hierbo genoem.

73 *Wills: a Practical Guide* (2011) 1.

74 Vgl Havenga "Murder for Insurance: Policy pays out 'life'" 2006 *Fundamina* 1-20 bespreek die toename in hierdie soort misdade met versekeringsbedrog.

hand van ware verhale, staaltjies en bekende misdade bespreek, die lig laat sien het nie. Ter aanvang verduidelik hy in die voorwoord: [75]

> *A would-be beneficiary from a deceased estate is not confined to ending the life of the distributor of largess in order to attain the desired end. There is the will itself. If need be, it can be forged or doctored; if there be one that a malefactor considers to be undesirable, it could be destroyed or hidden, or a later one could be forged.*[76]

Testamente word dikwels aangeveg deur ontevrede familie en vriende wat angstig is om te erf en wil seker maak dat hulle genoeg gaan erf.[77] Indien die testament nie iemand begunstig soos wat hy gehoop het sou gebeur nie, kan die versoeking ontstaan om die testament by die dood van die testateur te vernietig, te vervals of te verberg.[78] Fourie[79] waarsku dat familielede of 'n beswaarde persoon wat onterf is, dikwels die geldigheid van 'n erflater se testament betwis.[80] Hy wys daarop dat dit baie keer die geval met tweede huwelike en die botsende belange van die stiefma of stiefpa, en die kinders uit 'n vorige huwelik, is: Volgens hom gaan so 'n situasie gewoonlik gepaard met oor-en-weer beskuldigings van bedrog, onbehoorlike beïnvloeding en gierigheid. Aan die een kant is daar die belange van 'n langslewende gade en dié se behoefte om na omgesien te word, en aan die ander kant, die ontnugtering van die familie dat hul erfporsies nou daarmee heen is.[81]

Die omvang van probleme wat voorkom word geïllustreer in *Marais v Botha*[82] waar daar na sy dood 'n dispuut ontstaan oor die oorledene se

---

75 Kahn (2003) viii.

76 Kahn (2003) 147 ev bespreek onder die opskrifte "*Dirty work with a will*" tientalle sake en misdade wat hy klassifiseer onder die subopskrifte "*Failure to destroy a will; Concealment of a will; Doctored will*" en "*Forgery and falsification of a will*" probleme met testamente.

77 Vir geskille tussen familie sien oa *Reichman v Reichman* [2011] ZAGPJHP par 9; *Theart v Scheibert* [2012] ZASCA 131; *Levin v Levin* [2011] ZASCA 114; *Blom v Brown* 2011 3 All SA 223 (HHA)*; Taylor v Taylor* 2012 3 SA 219 (OK) en *Lipchick v Master of the High Court* [2011] ZAGPJHC 49 waar in par 1 opgemerk word: "*There has been tragic fraud amongst relations which played itself out in this case*". Vgl ook die bespreking hieronder en De Waal & Schoeman-Malan 39-45; King *Personal Finance* 2012-aanlyn; Hamill 2009 *Estate Planning*-aanlyn; Kahn (2003) 147 ev; Duhaime *Legal Resources* 2007-aanlyn*:* Gausden & King 1 ev.

78 Sien die opmerkings in *Marais v Botha* [2008] ZAWCHC 111; *Haribans v Haribans* [2011] ZAKZPHC 46 par 41. Sien ook Kahn (1984) 138: "*A fraud may be aware of the fact that an earlier will in his favour is still in existence but has been revoked by a later will not so favourable to him or leaving him nothing; if the circumstances are propitious he could destroy the later will and get away with it*."

79 (2012) die Hoofuitvoerendebeampte van PSG Konsultante.

80 Vgl die feite in *Reichman v Reichman* [2011] ZAGPJHP; *Theart v Scheibert* [2012] ZASCA 131; *Yokwana v Yokwana* [2013] ZAWCHC 22 par 7-8; *Webster v The Master* 1996 1 SA 34 (D); *Blom v Brown* 2011 3 All SA 223 (HHA); *Smith v Sampson* [2013] ZAWCHC 11; *Taylor v Taylor* 2012 3 SA 219 (OK).

81 Sien die feite in *Ex parte Porter* 2010 5 SA 546 (WK); *Levin v Levin* [2011] ZASCA 114; *Uys v Uys* [2008] ZANCHC 30; *Pienaar v Meester*; *Raubenheimer v Raubenheimer* 2012 5 SA 290 (HHA).

testament. Die oorledene se vrou is aangekla van sameswering tot moord op die oorledene. Die kinders van die oorledene het aangevoer dat die oorledene voor sy dood te kenne gegee het dat hy 'n nuwe testament verly het, aangesien hy vermoed het dat sy vrou hom probeer vergiftig. Die latere testament kon nie gevind word nie met die gevolg dat dit vermoedlik verberg of vernietig is. Die vrou doen aansoek dat 'n vroeëre gesamentlike testament geldig verklaar word. Die hof is nie oortuig dat daar nie wel 'n latere testamente is nie. Die saak *Lipchick v Master of the High Court*[83] is 'n verdere voorbeeld van verbittering en hartseer wat onderling by familielede oor die testament van die oorledene kan ontstaan. 'n Kodisil tot die testatrise se testament, wat sy self betwis, het al voor haar dood te voorskyn gekom. Sy ontken dat sy dit self geteken het. Dit blyk dat haar vooroorlede seun die kodisil (namens haar) opgestel het. Na haar dood is 'n testament gevind wat sy self opgestel het en waarin sy dié seun se vrou en kinders onterf. Die bitter twis het gegaan oor die seun se vrou en kinders wat deur die testatrise onterf is.[84]

Gerieflikheidshalwe kan vir doeleindes van die bespreking ten aansien van bestrede testamente, wat hulle by die dood van die testateur kan voordoen, onderskei word tussen gevalle waar:

(i) daar *'n testament gevind is*, maar die geldigheid daarvan betwis word;[85]

(ii) daar *nie 'n testament gevind is nie*, maar beweer word dat daar wel 'n testament was wat nie opgespoor kan word nie;[86]

(iii) 'n kombinasie van bogenoemde twee.[87]

Vir doeleindes van hierdie artikel sal elkeen van die drie gevalle in afsonderlike paragrawe bespreek word en nie as onderafdelings nie.

5 Waar 'n Testament Gevind is, Maar die Geldigheid Daarvan word Betwis

Twee moontlikhede kan hulle voordoen ten aansien van gevonde testamente:

82 [2008] ZAWCHC 111 par 1-2. 'n A 2(3)-aansoek word deur die vrou gebring om die vroeëre gesamentlike testament ingevolge waarvan sy erf te kondoneer. Sy dien tans gevangenisstraf uit vir sy moord. Die moordsaak is gerapporteer as *S v Marais* 2010 2 SACR 606 (KH).

83 [2011] ZAGPJHC 49.

84 Vgl ook *Harlow v Becker* 1998 4 SA 639 (D) waar die testatrise haar bates aan haar dokter nalaat en bevoordelings aan haar seun, kleinkinders en agterkleinkinders herroep. Sien verder *Thirion v Die Meester* 2001 4 SA 1078 (T) waar die ouers onterf en 'n minnares aangewys is as enigste erfgenaam; *Scott v Master of the High Court, Bloemfontein* [2012] ZAFSHC 190 waar die een seun onterf is en *Mabase v Dlamini* [2007] ZAGPHC 199 waar die vader van oorledene en die vrou met wie hy saambly gebly het stry oor die geldigheid van 'n dokument.

85 Par 5.

86 Par 6.

87 Par 7.

(a) Indien 'n testament gevind is, maar daar word aangevoer dat dit ongeldig is as gevolg daarvan dat die testament verval het, of daar bewerings van bedrog is, sal daar op 'n oorwig van waarskynlikhede bewys moet word dat die testament nie die van die testateur is nie.[88] Die hof se bevinding oor die geldigheid van die testament sal bepaal of daar aan die testament uitvoering gegee word of nie.[89] By bedrog of vervalsing word daar dikwels van handskrifdeskundiges gebruik gemaak om aan te toon dat die handtekening nie dié van die oorledene is nie.[90]

(b) Dit gebeur ook dat daar wel 'n testament gevind word, maar omdat die begunstigdes nie van die testament geweet het nie, of ongelukkig is met die uitkoms van die testament, word die geldigheid daarvan betwis.[91] Bewerings van onbehoorlike beïnvloeding, geestesongesteldheid, ensovoorts word dan aangevoer om die geldigheid van die latere testament aan te veg.[92]

5 1 Vervalste Testamente

Alhoewel die Boedelwet strafbepalings bevat wat handel met vervalsing van testamente, word artikel 102 van dié Wet nie dikwels aangewend om persone vir die vervalsing, diefstal, opsetlike vernietiging, verberging, of beskadiging van 'n testament te vervolg nie.[93] Uit gerapporteerde gevalle blyk dit dat 'n klag van bedrog gewoonlik eerder saam met ander strafregtelike klagte gelê word.[94] Die enkele sake waar artikel 102 van die Boedelwet wel toegepas is, is *S v Van Zyl*[95] waar die beskuldigde aangekla is daarvan dat hy die testateur se testament vervals het om homself as begunstigde aan te wys. Hy is skuldig bevind en tot drie jaar gevangenisstraf gevonnis waarvan 'n gedeelte opgeskort is. In *Pillay v Nagan*[96] het een van die intestate erfgename na die dood van die oorledene 'n testament vervals en te voorskyn gebring waarin hy

88 Corbett ea 117; Kahn (2003) 147 ev vir "*Forgery and falcification*".
89 Sien bespreking van die bewyslas hieronder par 8.
90 *Haribans v Haribans* [2011] ZAKZPHC 46 par 14; *De Reszke v Maras* 2006 2 SA 277 (HHA); *Katz v Katz* [2004] 4 All SA 545 (K); *Essop v Mustapha* & *Essop* 1988 4 SA 213 (D); Pienaar "Moord: Regter se testament vervals, beslis hof" *Beeld* (2013-01-31); Van der Westhuizen "Geen twyfel oor valse handtekening" *Beeld* (1997-09-18).
91 *Harlow v Becker* 1998 4 SA 639 (D); *Haribans v Haribans* [2011] ZAKZPHC 46*; Mabase v Dlamini* [2007] ZAGPHC 199; *Scott v Master of the High Court, Bloemfontein* [2012] ZAFSHC 190; *Smith v Sampson* [2013] ZAWCHC 11.
92 N 91 *supra*; Fourie *PSGKonsult*-aanlyn.
93 Vgl Sonnekus "Onwaardigheid vir Erfopvolging én die Versekerings-begunstigde" 2010 *TSAR* 175 177. Cronjé & Roos 87, waar dié skrywers daarop wys dat 'n testament vervals mag wees en dat dit onmoontlik kan wees om dit te bewys. Vgl ook Kahn (2003) "Forgery and falsification of a will" 152 ev.
94 Vgl bv Botha "Testament glo 'gekook'" *Volksblad* (2001-04-15) waar polisiebeamptes aangekla word van vervalsing deurdat hulle die oorledene se duimafdruk verkry het nadat hy 'n beroerte gehad het; Krige "Carmen Fortuin in die hof oor bedrog en moord op man" *Son*; "Atlantis-vrou in die hof oor bedrog en moord op man" *Son* (2007-01-23); Carstens "Sakeman vervals glo vrou se testament sodat hy kan erf" *Beeld* (2002-05-16) 6.
95 1985 3 SA 25 (A). Kahn (2003) 158 se bespreking van die saak.
96 2001 1 SA 410 (D); Cronjé & Roos 203; Kahn (2003) 159.

kwansuis bevoordeel is. Die hof het bevind dat aangesien hy die testament vervals het, hy onwaardig is om te erf. In *Rex v Foreman*[97] is 'n erfgenaam aangekla van bedrog waar hy twee testamente vervals het.[98] Die hof bevind dat albei die testament ongeldig is en dat 'n testament deur die bank opgestel geldig is. In die *Maidstone Crown Court* in Engeland in die saak *R v Avery* verskyn ses persone op aanklag van vervalsing van 'n testament. Na die oorledene verdrink het voer haar broer aan dat hy die enigste erfgenaam is ingevolge 'n testament wat hy in sy besit het. Die broer beweer haar voormalige man en kinders het 'n latere testament vervals. Dit blyk egter dat die broer al vantevore sy vader se testament vervals het. Dis staat laat vaar die klag van bedrog teen die oorledene se man en kinders.[99]

Bewerings van bedrog en vervalsing van 'n testament het onlangs weer ter sprake gekom na die moord op wyle waarnemende-regter Patrick Maqubela in 2010.[100] Sy vrou en 'n mede-beskuldigde staan ten tyde van die skryf hiervan tereg op 'n klag van moord op die waarnemende-regter. Sy word verder aangekla van bedrog na sy na bewering die testament vervals het.[101] Daar is aanvanklik gemeen dat die regter intestaat gesterf het waarna sy vrou as eksekuteur datief aangestel is. Nadat sy van moord aangekla is, is sy as eksekuteur onthef. 'n Klaarblyklike vervalste testament, ingevolge waarvan sy en haar drie dogters sou erf, tot uitsluiting van die oorledene se kinders uit 'n vorige huwelik, is by die meester ingehandig. Die testament is reeds deur die hof ongeldig verklaar.[102]

---

97 1952 1 SA 423 (SC). Vgl verder Eric Stafford *Natal Sunday Star* (1993-10-22) 1 en (1993-10-24) 5 en Du Plooy-Gildenhuys & Blackbeard 1999 *THRHR* 148.

98 Vgl ook *Tshona v Wauchope* 1908 EDC 32 vir vervalsing. Kahn (2003) 152-174 bespreek oa "Isadore Liondore's will – *the charge of forging and uttering*" en "The Lituanian will – *an unnecessary piece of play-acting*". In die eerste geval het die weduwee van 'n welgestelde testateur en mede-erfgenaam van die boedel daarop aangedring dat haar man 'n latere testament gemaak het waarin sy as enigste erfgenaam aangewys is. In die tweede geval het die weduwee ook 'n nuwe testament in 'n verseëlde koevert gevind. Albei testamente het geblyk vervals te wees.

99 "Will Forgery Claim Colapse" *Contesting Wills* (2012-02-07) beskikbaar by http://www.contesting-wills.co.uk/news-articles/will-forgery-claim-colapses.html (besoek 2013-04-11) "*The deceased, Karen Griffin, drowned in the River Medway in November 2009. She was claimed to have left a Will made in August 2004 under which her entire estate was left to her brother, David Banks. The deceased's ex husband, Bernard Griffin, her children Daniel, Simon and Emma Griffin and partner, Raymond Avery, together with her neighbour, Alan McNicol were subsequently alleged to have forged another Will under which Simon Griffin was the sole beneficiary.*"

100 *S v TS Maqubela* (WKH 2011-11-12. Beskikbaar by Court rolls at http://www.saflii.org/blog/?page_id=134). Vgl Mfazwe "Regtersvrou ook verkla van bedrog" en "Twee erf niks 'vervalste testament' steeds onder die loep" *Die Burger* (2011-11-02) 8.

101 Breytenbach "Hof weier Maqubela se ontslag-aansoek" *Die Burger* (2013-03-07); Mackay "Maqubela trial told of meeting" *Sowetan* (2013-04-10).

102 De Wee "Regtersvrou ook verkla van bedrog" *Die Burger* (2011-11-02) 8; Pienaar "Moord: Regter se testament vervals, beslis hof" *Beeld* (2013-01-31).

In die laat 1990s het die saak van moord en bedrog teen 'n boer van Lydenburg, groot opslae in die pers gemaak.[103] Hy word vyf jaar na sy vrou wreed vermoor is, en hy miljoene aan polisuitbetalings geëis het, aangekla en skuldig bevind aan moord en vervalsing van haar testament.[104]

5 2 Gevalle Waar 'n Testament Gevind is, Maar Waarna die Geldigheid Betwis word

'n Tweede geval waar daar wel 'n testament gevind word, waarvan begunstigdes onbewus was, wat aanleiding kan gee tot ernstige geskille, is die ontdekking van 'n latere testament. So 'n testament, wat na die dood van die testateur te voorskyn kom, kan aanleiding gee tot groot ongelukkigheid onder potensiële erfgename.[105] Indien naasbestaandes nie bevoordeel word nie, of nie soveel erf soos hulle gehoop het nie,[106] word die geldigheid van die testament gewoonlik betwis.[107]

In *Levin v Levin*[108] het die oorledene negentien verskillende testamente verly. Sy het haar familie probeer beheer deur gereeld haar testament te verander en die erfporsies te vergroot of te verklein en begunstigdes in te sluit of uit te sluit, afhangende van wie haar op daardie tydstip goedgesind was of nie. Sy het die laaste testament in die geheim opgestel (al die vorige testamente is deur die bank opgestel) om die voortdurende onmin tussen haar oorlewende kinders en kleinkinders te

103 Die verloop van die strafsaak is deur die pers gedek: Van der Westhuizen "Miljoenêr-boer se voorman getuig oor dié testament 'Hy het my vertel hy't dit vervals'" *Beeld* (1997-08-01) 3. Daar is beweer dat hy klaarblyklik iemand gekry het om namens haar 'n testament te teken.

104 Van der Westhuizen "Testament van miljoener-boer. Wie't geteken? Getuies erken vermoorde nie" *Beeld* (1997-07-31) 1 en "Du Toit het reuse-skuld só afgelos. Erfgeld laat miljoenêr glo kop bo water hou" *Beeld* (1997-08-22). Du Toit het die versekeringsmaatskappy, wat nie die polisse wat hy op sy vrou se lewe uitgeneem het, wou uitbetaal nie voor die moordklag gedagvaar: sien *Du Toit v Standard General Insurance Co Ltd* 1994 1 SA 682 (W). In *Haribans v Haribans* [2011] ZAKZPHC 46 par 41 merk die hof op dat 'n afskrif van die testament waarskynlik vervals is.

105 Die feite in *Harlow v Becker* 1998 4 SA 639 (D); *De Reszke v Maras* [2006] 2 All SA 115 (SCA)*; Haribans v Haribans* [2011] ZAKZPHC 46*; Mabase v Dlamini* [2007] ZAGPHC 199; *Scott v Master of the High Court, Bloemfontein* [2012] ZAFSHC 190; *Smith v Sampson* [2013] ZAWCHC 11.

106 Kahn (2003) 147 ev; Heckenberg 2012 *Find Law Australia*. Sien ook *Harlow v Becker* 1998 4 SA 639 (D) en *Lipchick v Master of the High Court* [2011] ZAGPJHC 49 waar die testamente (wat nie behoorlik verly was nie) betwis word deur van die oorledenes se kleinkinders wat in die testamente onterf is, en *Leksekga v The Master* 1995 4 SA 731 (W).

107 *Lipchick v Master of the High Court* [2011] ZAGPJHC 49 par 8; *Haribans v Haribans* [2011] ZAKZPHC 46; *Mabase v Dlamini* [2007] ZAGPHC 199; *Scott v Master of the High Court, Bloemfontein* [2012] ZAFSHC 190; *Smith v Sampson* [2013] ZAWCHC 11. Vir Geestesongesondheid en onbehoorlike beïnvloeding sien n 91 *supra* en spesifiek *Katz v Katz* [2011] ZAGPJHC 224; *De Reszke v Maras* [2006] 2 All SA 115 (SCA).

108 [2011] ZASCA 114 par 9.

probeer besleg.[109] By haar dood op 107-jarige ouderdom is die erfgename steeds ontevrede en betwis die geldigheid van die laaste testament. In *Lipchick v Master*[110] het die testatrise 'n handgeskrewe testament waarin sy sekere van haar naasbestaandes onterf en waarvan haar erfgename onbewus was, aan haar jare lange vriend oorhandig. Na haar dood word die geldigheid van dié testament, wat nie deur getuies onderteken was nie, deur 'n ongelukkige skoondogter en kleinkinders betwis.

Blom v Brown[111] is 'n verdere saak waar 'n gebrek aan kennis oor 'n latere testament (wat by die dood van die testateur gevind is) aan die kant van 'n erfgenaam tot 'n geskil en litigasie lei. 'n Familietwis ontstaan toe dit na die dood van die testateur blyk dat hy 'n testament, waarin sy tweede vrou as enigste erfgenaam aangewys is, verly het. Sy twee dogters uit sy eerste huwelik (wat onterf is) veg die geldigheid van die latere testament aan op grond daarvan dat die tweede vrou die testament geskryf het. Die hof bevind die tweede testment is geldig verly. In *Thakar v Naran*[112] is 'n tweede "nuwe" testament eers sewe jaar na die dood van die oorledene gevind. Die oorledene pleeg selfmoord in 1978 deur sianied te drink. In 1986 word 'n testament deur sy seun gevind wat hy klaarblyklik die dag van sy selfmoord verly het.[113] Die geldigheid van 'n "nuwe testament" word ook in *Ex parte Warren*[114] betwis waar 'n vriendin van die oorledene die nuwe testament aanveg waarin sy onterf is. Die oorledene en haar vriendin, albei oujongnooiens, het saamgewoon en mekaar oor en weer bevoordeel in hulle testamente. Die oorledene word op 'n stadium deur 'n familielid geneem na 'n versorgingseenheid vir drankmisbruikers. Sy verly daarna 'n testament waarin sy die oujongnooivriendin onterf en haar moeder as erfgenaam van 'n gedeelte van haar boedel instel. Dié vriendin beweer dat die oorledene deur haar familie beïnvloed is om haar te onterf.

109 Vgl ook De Bruin "Moeilike tannie lag laaste in haar groot testament" *Beeld* (2011-6-6) 8 waar die testatrise beskryf word as 'n skatryk, maar "baie moeilike, dominerende, manipulerende, suinige en agterdogtige" weduwee wat in 2002 oorlede is. Die "nuwe" testament is deur die hof aanvaar as die laaste geldige testament.

110 [2011] ZAGPJHC 49.

111 [2011] 3 All SA 223 (HHA).

112 1993 4 SA 665 (N). Vir nog nuusberigte waar latere nuwe testament te voorskyn kom sien Potgieter "Hy onterf hul, nou is as weg" *Beeld* (2012-02-24); "Drie weier om oor boedel te praat" *Die Burger* (2012-02-25) waar die kinders van die erflater onterf word en hy sy prokureur en sekretaresse bevoordeel. Vgl ook Jansen "Hof toe ná pa hulle onterf" *Rapport* (2012-12-09) en Potgieter "Kinders hof toe oor R110 m. nadat boer hulle onterf" *Die Burger* (2012-02-22).

113 Vgl ook *Thirion v Die Meester*.

114 1955 4 SA 326 (W). Vgl ook *Bosch v Nell* 1992 3 SA 600 (T) waar die vrou van die testateur aanvanklik as die enigste erfgenaam en eksekuteur aangewys. In 'n tweede testament, waarvan sy onbewus was, is die boedel tussen haar en drie ander persone verdeel en 'n ander eksekuteur aangestel. Sy betwis die geldigheid van die tweede testament op grond van die feit dat die erflater nie die testament in teenwoordigheid van twee getuies geteken het nie.

Gevalle waar 'n testament *nie by die dood van die erflater gevind is nie* maar beweer word dat daar wel 'n testament was, word in paragrawe 6 en 7 in Deel 2 bespreek.

Cosmetic surgery and responsible patient selection – does a legal duty to screen patients exist?

Hanneke Verwey
LLB LLM
Academic Associate, Department of Private Law, University of Pretoria

Pieter Carstens
BLC LLB LLD
Professor and Head of the Department of Public Law, University of Pretoria

OPSOMMING

Kosmetiese Chirurgie en Verantwoordelike Pasiëntkeuse – Bestaan daar 'n Regsplig om Pasiënte te Evalueer?

Daar is sekere unieke uitdagings wat pasiëntkeuse vir kosmetiese chirurgie betref. Die uitdaging wat kosmetiese chirurge in die gesig staar is hoe om reeds voor chirurgiese ingryping pasiënte te identifiseer wat 'n swak resultaat in terme van sielkundige aanpassing en psigososiale funksionering sal hê, ten spyte van 'n tegnies aanvaarbare resultaat. Aangesien sielkundige faktore onderliggend is aan die meeste versoeke vir kosmetiese chirurgie, word daar aan die hand gedoen dat kosmetiese chirurge 'n regsplig het om bewus te wees van algemeen erkende psigiatriese toestande en simptome. Kosmetiese chirurge, meer so as die meeste ander spesialiste, moet aandagtige luisteraars wees met kliniese vernuf wat strek buite die tipiese grense van medisyne en siekte. Die deliktuele, strafregtelike en kontraktuele aanspreeklikheid van 'n kosmetiese chirurg wat opereer op 'n sielkundig onstabiele pasiënt word in besonderhede bespreek. Die bespreking vind plaas aan die hand van 'n Amerikaanse saak, die enigste saak van sy soort ter wêreld, waar die vraag of 'n pasiënt wat ly aan liggaamsdismorfiese versteuring geldige toestemming tot kosmetiese chirurgie kan gee, aangespreek is.

1 Introduction

Medical techniques and technology are increasingly being used for purposes that seemingly deviate from the traditional goals of medicine. Medical techniques and technology are often implemented, not to prevent or cure illness, but to fulfil a patient's personal, individual and ostensibly non-medical wishes.[1] These wishes are often aimed at improving certain human characteristics beyond their normal healthy state.[2] A prime example of wish-fulfilling medicine is cosmetic surgery. The pursuit of beauty by means of cosmetic surgery is big business in modern societies and South Africa is catching up very fast in this

1 Buyx "Be Careful What You Wish For? Theoretical and Ethical Aspects of Wish-Fulfilling Medicine" 2008 *Medicine, Healthcare and Philosophy* 134.

2 Bordo "Material girl: the Effacements of Postmodern Culture" 1990 *Mich Qly R* 657; Gimlin "Cosmetic Surgery: Beauty as Commodity" 2000 *Qualitative Sociology* 80.

particular area. With the rise of cosmetic surgery, the contemporary body, instead of being a dysfunctional object requiring medical interventions, has become a primary symbol of identity and a commodity, not unlike "a car, a refrigerator, a house, which can be continuously upgraded and modified in accordance with new interests and greater resources".[3] Cosmetic surgery has in fact become a "modern body custom".[4] As exciting as this might be to some, certain legal and ethical issues must not be overlooked. The fact remains that cosmetic surgery involves the performance of very invasive surgical operations on otherwise healthy individuals for the sake of improving appearance. Miller *et al* describe cosmetic surgery as "a most unusual medical practice" and state[5] that:

> [i]nvasive surgical operations performed on healthy bodies for the sake of improving appearance lie far outside the core domain of medicine as a profession dedicated to saving lives, healing, and promoting health.

Due to the ethically ambiguous nature of cosmetic surgery, it is submitted that certain safeguards must be put in place in order to prevent ethical abuses. In the everyday practice of a cosmetic surgeon, at least some of these safeguards need to address the process of informed consent and patient selection.

2 Why the Relationship between Cosmetic Surgeons and their Patients Differs from the Conventional Doctor-Patient Relationship

The performance of cosmetic surgery necessitates a degree of ethical conduct on the part of the cosmetic surgeon that surpasses the level of ethical conduct normally required between a physician and patient as the relationship between a cosmetic surgeon and a patient differs from the traditional physician-patient relationship.[6] This is essentially due to the distinction, albeit tenuous, between elective and non-elective forms of medical treatment. Distinguishing between elective and non-elective medical treatments is difficult, but cosmetic surgery is usually elective in the sense that cosmetic surgery is opted for by a patient more freely and less for reasons of medical necessity in the narrow sense of the word.[7]

---

3 Finkelstein *The Fashioned Self* (1991) 87; Gimlin 2000 *Qualitative Sociology* 80; Adams "Motivational Narratives and Assessments of the Body after Cosmetic Surgery" 2010 *Qualitative Health Research* 757.

4 Sullivan *Cosmetic Surgery, the Cutting Edge of Commercial Medicine in America* (2000) 10.

5 Miller *et al* "Cosmetic Surgery and the Internal Morality of Medicine" 2000 *Camb Qly Healthcare Ethics* 353.

6 Atiyeh *et al* "Aesthetic/Cosmetic Surgery and Ethical Challenges" 2008 *J Aesthetic and Plastic Surgery* 830.

7 Healy "Duties of Disclosure and the Elective Patient: a Case for Informed Consent" *1998 Medico-Legal J Ire* 26; Nugent "Cosmetic surgery on Patients with Body Dysmorphic Disorder: the Medical, Legal and Ethical

In some countries the courts have been hesitant to accept a distinction between elective and therapeutic or non-elective procedures, as all operations are elective in the sense that the patient always has a choice whether or not to undergo the procedure.[8] What these courts have not taken into consideration is the fact that there is a very real distinction between situations where the patient has very little choice but to undergo the procedure, as the treatment is indicated as the best or only option, and situations where the patient can comparatively afford not to undergo the procedure. This viewpoint was expressed quite eloquently by McCarthy J in the Irish case of *Walsh v Family Planning Services Ltd* when he held that:

> All surgery, in a sense, is elective although the election may have to be implied from the circumstances rather than determined as express ... A patient's condition may be such as to demand surgical intervention as the only hope for survival. Such may be called non-elective surgery. The patient given the choice between enduring pain and having limb replacement surgery or fusion surgery may technically be electing as between pain and the surgery but the election may be more apparent than real. An extreme of elective surgery would be what is purely cosmetic – simply to improve the natural appearance rather than to remedy the physical results of injury or disease. Even it may have an element of quasi-medical care because of the psychological reaction of the patient to personal appearance. A like argument may be advanced in respect of contraceptive surgery, male or female. Such surgery does not have a direct effect on the health or wellbeing of the patient nor in prolongation of life; it may alleviate marital stress or other domestic pressure and in that sense be therapeutic. Essentially, however, it is for the improvement of the sex life of the couple concerned.[9]

Cosmetic surgery, as an example of elective surgery, is a treatment which comparatively, the patient can afford not to undergo.[10] Conventionally, a patient experiencing specific symptoms seeks help from a physician and the physician makes a subsequent diagnosis based on objective scientific knowledge.[11] The diagnosis is followed by the performance of a suitable treatment, provided of course that the physician has obtained the patient's informed consent to the administration of such treatment.[12] Conversely, cosmetic surgery patients generally have no symptoms and therefore a resultant diagnosis

---

Implications" http://works.bepress.com/cgi/viewcontent.cgi?article = 1004& context = Kristen_nugent (accessed 2012-02-14); Buyx 2008 *Medicine, Healthcare and Philosophy* 135.

8 *Sidaway v Bethlem Royal Hospital Governors* 1985 1 ALL ER 643 (HL); *Gold v Haringey Health Authority* 1987 2 ALL ER 888 (CA).

9 *Walsh v Family Planning Services* 1992 1 IR 496 517-518.

10 Healy 1998 *Medico-Legal J Ire* 27.

11 Nordenfelt "The Concepts of Health and Illness Revisited" 2007 *Medicine, Health Care and Philosophy* (*Med Health Care Philos*) 8; Buyx 2008 *Medicine, Healthcare and Philosophy* 135.

12 *Ibid.*

is impossible.[13] When performing cosmetic surgery, cosmetic surgeons are subjecting otherwise perfectly healthy individuals to medical risks, side effects and complications for benefits that are, arguably, non-medical.[14] The treatment selection is determined, or at the very least guided, by the patient's wishes.[15] The patient chooses to have cosmetic surgery, rather than the surgery being an absolute necessity, therefore the decision for surgery is a joint process.[16] Communication between the cosmetic surgeon and the patient takes place on a different level as the patient typically expects to relate more democratically with the cosmetic surgeon.[17] Positions of interaction are therefore uniquely different for both the patient and the cosmetic surgeon.[18] In the case of therapeutic or non-elective operations the patient is often reluctant to consent to surgery and must even be persuaded by the physician, whereas the cosmetic surgery patient requests the operation and sometimes actually talks the cosmetic surgeon into performing it.[19] The cosmetic surgeon does not play a crucial role in determining the course of treatment and primarily acts as a source of information to the patient.[20]

Furthermore, when aesthetics is involved, the success of the treatment is entirely dependent upon the patient's subjective opinion.[21] More so than in the case of non-elective or therapeutic surgery, psychological factors are particularly relevant in the case of cosmetic surgery.[22] The motivation for cosmetic surgery differs from therapeutic or non-elective surgery and is often overlooked by the cosmetic surgeon or disguised by the patient.[23] The patient may appear to be very well adjusted, but it

13 Buyx 2008 *Medicine, Healthcare and Philosophy* 135; Heyes & Jones "Cosmetic Surgery in the Age of Gender" in *Cosmetic Surgery: a Feminist Primer* (ed Heyes) (2009) 5.

14 Devereaux "Cosmetic Surgery" in *Medical Enhancement and Posthumanity* (eds Gordijn & Chadwick) (2009) 163; Buyx 2008 *Medicine, Healthcare and Philosophy* 134.

15 Simon "Physician's Duty to Screen Patients for Elective Surgery" 1978 *Ariz LR* 670; Buyx 2008 *Medicine, Healthcare and Philosophy* 135; Heyes & Jones 5.

16 Wright "Management of Patient Dissatisfaction with Results of Cosmetic Procedures" 1980 *Arch Otolarygol Head Neck Surg* 466; Heyes & Jones 5.

17 Wright 1980 *Arch Otolarygol Head Neck Surg* 467; Ezekiel & Emanuel "Four Models of the Physician-Patient Relationship" 1992 *J Am Med Ass* 2221.

18 Wright 1980 *Arch Otolarygol Head Neck Surg* 467; Buyx 2008 *Medicine, Healthcare and Philosophy* 134.

19 Wright & Wright "A Psychological Study of Patients Undergoing Cosmetic Surgery" 1975 *Arch Otolarygol Head Neck Surg* 145.

20 Buyx 2008 *Medicine, Healthcare and Philosophy* 135.

21 Preminger & Fins "Plastic Surgery, Aesthetics and Medical Professionalism: Beauty and the Eye of the Beholder" 2009 *Ann Plast Surg* 342; Ericksen & Billick "Psychiatric Issues in Cosmetic Plastic Surgery" 2012 *Psychiatric Qly* 2; Francis "Informed Consent in Body Dysmorphic Disorder" http://www.medscape.com/viewarticle/758800_4 (accessed 2013-02-20).

22 Wright 1980 *Arch Otolarygol Head Neck Surg* 467; Sarwer *et al* "Psychological Investigations in Cosmetic Surgery: a Look Back and a Look Ahead" 1998 *Plastic and Reconstructive Surgery* 1137; Ericksen & Billick 2012 *Psychiatric Qly* 2.

23 Wright & Wright 1975 *Arch Otolarygol Head Neck Surg* 145.

must be kept in mind that cosmetic procedures will always involve the patient's psyche.[24] The cosmetic surgery patient has the luxury of indulging in his or her own personal whims, wishes, and unrealistic expectations.[25] The cosmetic surgeon must therefore be cognisant of underlying psychological manifestations and take into consideration that behind every request for cosmetic change, there is always the desire for an improved self-image or self-concept.

Changes in the public's attitude toward the physician's role are particularly prominent in cosmetic surgery because of the differences between cosmetic surgery and other areas of medicine. We live in an era where physicians are seen as health providers and patients as health consumers within a larger health industry. According to Wright, technological advancements have widened the sphere of modern medicine, but it also tends to objectify medicine and mislead the public.[26] An inherent danger is that medicine can be oversold and patients may unconsciously try to contract with their physicians for a product or have unrealistic expectations of success.[27] Then, when the patients find that despite the wonders of modern medicine they cannot buy miracles, they may still want to see the physician as an all-knowing healer.[28] Modern-day physicians, particularly cosmetic surgeons, therefore find themselves confronted with patients who apparently understand the capabilities of modern medicine and who want to have more say in the physician-patient relationship, yet they do not want to abandon the traditional concept of the physician's role.[29]

Cosmetic surgeons, more so than any other medical specialists, also face inherent conflicts of interests as performing cosmetic surgery and other cosmetic procedures is a lucrative venture, particularly as it is often a market of pathological repeat customers.[30] To some extent all physicians face a variation of this problem, namely that their livelihoods depend on performing the interventions they recommend.[31] However, economic self-interest is far less blatant when a general surgeon insists that a sick patient have an appendectomy, even if he or she stands to profit from the procedure, as opposed to when a cosmetic surgeon

24 Wright 1980 *Arch Otolarygol Head Neck Surg* 467; Sarwer *et al* 1998 *Plastic and Reconstructive Surgery* 1137.
25 Wright 1980 *Arch Otolarygol Head Neck Surg* 467; Ringel "The Morality of Cosmetic Surgery for Aging" 1998 *Arch Derm* 430.
26 Wright 1980 *Arch Otolarygol Head Neck Surg* 466.
27 *Ibid.*
28 Wright 1980 *Arch Otolarygol Head Neck Surg* 466; Bordo "Twenty Years in the Twilight Zone" in *Cosmetic Surgery: a Feminist Primer* (ed Heyes) (2009) 28.
29 Wright 1980 *Arch Otolarygol Head Neck Surg* 466.
30 Cantor "Cosmetic Dermatology and Physician's Ethical Obligations: More Than Just Hope in a Jar" 2005 *Seminars in Cutaneous Medicine and Surgery* 155; Atiyeh *et al* 2008 *J Aesthetic and Plastic Surgery* 834 Heyes & Jones 5.
31 Cantor 2005 *Seminars in Cutaneous Medicine and Surgery* 155; Atiyeh *et al* 2008 *J Aesthetic and Plastic Surgery* 834.

suggests a patient undergo some extra liposuction along with their abdominoplasty or a chin augmentation along with their rhinoplasty.[32]

3 Cosmetic Surgery and Patient Selection – the Challenges

There are unique challenges to cosmetic surgery patient selection. Considering the large numbers of individuals who choose to undergo cosmetic surgery, it is probable that all of the major psychiatric diagnoses occur in this population.[33] However, certain disorders, particularly those with a body image component, may be more prevalent in cosmetic surgery patients and may contraindicate surgery.[34] Furthermore, studies have shown there to be a strong correlation between preoperative psychological problems and adverse postoperative effects such as depression, general unhappiness with the surgical results, suicidal thoughts, feelings of anger and resentment towards the cosmetic surgeon and the onset of psychotic episodes.[35] Studies have shown that factors such as youthfulness, unrealistic expectations, previous unsatisfactory cosmetic surgery, a disproportionate concern over a minimal deformity, motivation based on relationship issues and a history of depression, anxiety, body dysmorphic disorder, thought disorders and eating disorders are all associated with poor psychological outcomes.[36] If surgery does not prove to be a solution to these problems, these types of patients might have severe adjustment traumas when they come to the realisation that their difficulties were caused by factors other than

32 *Ibid.*

33 Grossbart & Sarwer "Psychosocial Issues and Their Relevance to the Cosmetic Surgery Patient" 2003 *Seminars in Cutaneous Medicine and Surgery* 140; Crerand *et al* "Body Dysmorphic Disorder and Cosmetic Surgery" 2006 *Plastic and Reconstructive Surgery* 173; Newell "Informed Consent for Plastic Surgery. Does It Cut Deeply Enough?" 2011 *J Leg Med* 328.

34 Holder "Cosmetic Breast Surgery" 1972 *J Am Med Ass* 1102; Grossbart & Sarwer 2003 *Seminars in Cutaneous Medicine and Surgery* 140; Cantor 2005 *Seminars in Cutaneous Medicine and Surgery* 159; Crerand *et al* 2006 *Plastic and Reconstructive Surgery* 172.

35 Reich "Factors Influencing Patients Satisfaction With the Results of Aesthetic Plastic Surgery" 1975 *Plastic and Reconstructive Surgery* 5; Wright & Wright 1975 *Arch Otolarygol Head Neck Surg* 145; Simon 1978 *Ariz LR* 672; Cantor 2005 *Seminars in Cutaneous Medicine and Surgery* 159.

36 Goin *et al* "A Prospective Psychological Study of 50 Female Face-Lift Patients" 1980 *Plastic and Reconstructive Surgery* 436; Beale *et al* "Augmentation Mammoplasty: the Surgical and Psychological Effects of the Operation and Prediction of the Result" 1985 *Ann Plast Surg* 473; Grossbart & Sarwer 2003 *Seminars in Cutaneous Medicine and Surgery* 142; Honigman *et al* "A Review of Psychosocial Outcomes For Patients Seeking Cosmetic Surgery" 2004 *Journal of Plastic and Reconstructive Surgery (Plastic and Reconstructive Surgery)* 1229; Ericksen & Billick 2012 *Psychiatric Qly* 5; Nugent "Cosmetic Surgery on Patients With Body Dysmorphic Disorder: the Medical, Legal and Ethical Implications" http://works.bepress.com/cgi/viewcontent.cgi?article=1004&context=Kristen_nugent (accessed 2012-02-14)

their physical appearance.[37] When psychological problems are blamed solely on a cosmetic deficiency, removal of the deficiency my actually increase the severity of the problem and the cosmetic surgeon risks feeding the patient's pathology and violating the ethical principle of non-maleficence.[38] An emotional reaction such as this may lead to allegations that the cosmetic surgeon, not the patient, was at fault.[39] Patients like these are generally hard to please and can become litigation-minded quite easily.[40] Cosmetic surgeons are taught how to and when to operate, but being taught when not to operate is often discounted.[41] The challenge that cosmetic surgeons face is how to identify, prior to surgical intervention, those patients who may have a poor result in terms of psychological adjustment and psychosocial functioning in spite of a technically acceptable result.[42] As "appearance is so laminated to psychological cohesion"[43] and because psychological factors are at the root of most requests for cosmetic surgery, it is submitted that a legal duty on the part of the cosmetic surgeon to remain sensitive to widely recognised dangerous symptomology should exist.

4 The Psychiatric or Psychological Evaluation

It would of course be unreasonable to expect cosmetic surgeons to conduct extensive personality tests on their patients.[44] An actual psychiatric or psychological evaluation is impractical as a routine screening process.[45] Psychiatric or psychological evaluations are expensive, laborious and time-consuming.[46] They certainly represent an astute patient selector, but besides being impractical, psychiatric or psychological evaluations could also be unreliable if employed by the untrained, inexperienced interviewer.[47] A true psychiatric or psychological evaluation is also likely to put the real problem patients on

37 Holder 1972 *J Am Med Ass* 1102.
38 Holder 1972 *J Am Med Ass* 1102; Cantor 2005 *Seminars in Cutaneous Medicine and Surgery* 159.
39 Holder 1972 *J Am Med Ass* 1102.
40 Holder 1972 *J Am Med Ass* 1102; Nugent "Cosmetic Surgery on Patients With Body Dysmorphic Disorder: the Medical, Legal and Ethical Implications" http://works.bepress.com/cgi/viewcontent.cgi?article = 1004&context = Kristen_nugent (accessed 2012-02-14); Carstens & Pearmain *Foundational Principles of South African Medical Law* (2007) 708.
41 Sykes 2009 *Curr Opin Otolaryngol Head Neck Surg* 322.
42 Honigman *et al* 2004 *Plastic and Reconstructive Surgery* 1230; Sykes "Managing the Psychological Aspects of Plastic Surgery Patients" 2009 *Curr Opin Otolaryngol Head Neck Surg* 322.
43 Jefferson "The Psychiatric Assessment of Candidates For Cosmetic Surgery" 1976 *J Nat Med Ass* 411; Simon 1978 *Ariz LR* 685.
44 Nugent "Cosmetic Surgery on Patients With Body Dysmorphic Disorder: the Medical, Legal and Ethical Implications" http://works.bepress.com/cgi/viewcontent.cgi?article = 1004&context = Kristen_nugent (accessed 2012-02-14); Cantor 2005 *Seminars in Cutaneous Medicine and Surgery* 159.
45 Simon 1978 *Ariz LR* 680.
46 *Ibid.*
47 Wright & Wright 1975 *Arch Otolarygol Head Neck Surg* 150.

their guards as they will probably recognise what is going on and manipulate their way through the screening process.[48] According to Wright and Wright, a simplified, fairly stereotyped counselling routine is needed.[49] Such a routine would ideally encourage the patient to communicate forthrightly with his or her cosmetic surgeon and would safeguard even the less experienced cosmetic surgeons.[50]

Issues pertaining to the perceived physical flaw itself should be addressed.[51] Patients should articulate and describe their concern with their appearance in detail.[52] They should be questioned with regards to the length of time they have been concerned with the perceived deformity, the length of time they have been considering undergoing cosmetic surgery and the circumstances that occasioned the current consultation.[53] The cosmetic surgeon must determine whether the patient has undergone any previous cosmetic procedures and whether the patient was satisfied with the results of those surgeries. A history of previous surgeries, performed by different cosmetic surgeons, particularly if the patient was dissatisfied with most or all of these procedures, is reason for concern.[54] Any history of legal proceedings or apparent hostility toward previous cosmetic surgeons should also raise major concern.[55] Unrealistic expectations and improper motivations regarding the outcome of the procedure have also been shown to predict a poor psychological outcome.[56] The cosmetic surgeon should therefore assess the patient's expectations of both the proposed procedure and the desired outcome in cosmetic and personal terms.[57] The patient should be questioned to determine exactly what he or she desires, the realism of his or her expectations and his or her ability to accept imperfect results.[58] A distinction should be made between expectations regarding

48 Wright & Wright 1975 *Arch Otolarygol Head Neck Surg* 150; Simon 1978 *Ariz LR* 680.
49 Wright & Wright 1975 *Arch Otolarygol Head Neck Surg* 150.
50 *Ibid.*
51 Honigman *et al* 2004 *Plastic and Reconstructive Surgery* 1235.
52 Pruzinsky "Psychological Factors in Cosmetic Plastic Surgery: Recent Developments in Patient Care" 1993 *Plastic Surgical Nursing* 64; Honigman *et al* 2004 *Plastic and Reconstructive Surgery* 1235; Panfilov & Larkin *Cosmetic Surgery Today* (2005) 12.
53 Honigman *et al* 2004 *Plastic and Reconstructive Surgery* 1235.
54 Knorr "Feminine Loss of Identity in Rhinoplasty" 1972 *Arch Otolaryngol Head Neck Surg* 11, Goin *et al* 1980 *Plastic and Reconstructive Surgery* 436; Honigman *et al* 2004 *Plastic and Reconstructive Surgery* 1235.
55 Honigman *et al* 2004 *Plastic and Reconstructive Surgery* 1235.
56 Pruzinsky 1993 *Plastic Surgical Nursing* 64; Honigman *et al* 2004 *Plastic and Reconstructive Surgery* 1235; Sykes 2009 *Curr Opin Otolaryngol Head Neck Surg* 324; Panfilov & Larkin 12.
57 Aronoff "The Psychiatric Aspects of Rhinoplasty" in *Plastic and Reconstructive Surgery of the Face and Neck: Aesthetic Surgery* (eds Conley & Dickinson) (1972) 97; Wright & Wright "A Psychological Study of Patients Undergoing Cosmetic Surgery" 1975 *Arch Otolarygol Head Neck Surg* 150; Simon 1978 *Ariz LR* 673; Pruzinsky 1993 *Plastic Surgical Nursing* 64; Honigman *et al* 2004 *Plastic and Reconstructive Surgery* 1235.
58 Reich 1975 *Plastic and Reconstructive Surgery* 9; Simon 1978 *Ariz LR* 682; Panfilov & Larkin 12.

an improved self (for example an improved self-esteem) and expectations regarding enhanced external parameters (for example ending social isolation or getting a promotion at work).[59] Studies have shown the latter to be a reason for concern.[60] If the patient thinks that cosmetic surgery will provide a solution to social or interpersonal problems, the cosmetic surgeon should be wary of performing the procedure.[61] The cosmetic surgeon must carefully clarify what cosmetic surgery can and cannot accomplish.[62] If a patient refuses to enter into a mutual contract of responsibility with the cosmetic surgeon or does not view the surgery realistically, he or she may become a problem patient.[63] Simply asking a patient to state his or her motivation and expectations for the surgery may be insufficient. The mentally unstable patient may give misleading or inaccurate reasons for wanting the surgery and conceal his or her true, unhealthy motivations.[64] The reality is that many patients are far too sophisticated to reveal their true, unhealthy motivations during a simple diagnostic examination.[65] However, the cosmetic surgeon might be able to spot dangerous symptomology by paying attention to a patient's unrealistic responses to strategic questions.[66] Regardless of the patient's ability to camouflage underlying motivations, a cosmetic surgeon should at the very least make reasonable inquiries in order to identify danger signs widely recognised in the practice of cosmetic surgery.[67] Questions should focus on the decision-making process of the patient, family and marital relations, support shown by friends and family and the autonomy of the patient's decision.[68] The cosmetic surgeon must be wary of great enthusiasm, vague or disproportionate expectations, personal stress and paranoia.[69]

Studies have shown that a person with a history of depression, anxiety or a personality disorder is not an ideal candidate for cosmetic surgery in

59 Sarwer & Didie "Body Image in Cosmetic Surgical and Dermatological Practice" in *Disorders of Body Image* (eds Castle and Phillips) (2002) 37; Honigman *et al* 2004 *Plastic and Reconstructive Surgery* 1235.

60 Honigman *et al* 2004 *Plastic and Reconstructive Surgery* 1235 ; Panfilov & Larkin 12.

61 Honigman *et al* 2004 *Plastic and Reconstructive Surgery* 1235.

62 Wright & Wright 1975 *Arch Otolarygol Head Neck Surg* 150.

63 *Idem* 151.

64 Simon 1978 *Ariz LR* 671.

65 Wright & Wright 1975 *Arch Otolarygol Head Neck Surg* 146; Simon 1978 *Ariz LR* 671.

66 Wright & Wright 1975 *Arch Otolarygol Head Neck Surg* 148-50; Peterson & Topazian "Psychological Considerations in Corrective Maxillary and Midfacial Surgery" 1976 *J Oral Surg* 157; Simon 1978 *Ariz LR* 671.

67 Simon 1978 *Ariz LR* 671; Pitts-Taylor *Surgery Junkies: Wellness and Pathology in Cosmetic Culture* (2007) 131; Nugent "Cosmetic Surgery on Patients With Body Dysmorphic Disorder: the Medical, Legal and Ethical Implications" http://works.bepress.com/cgi/viewcontent.cgi?article = 1004& context = Kristen_nugent (accessed 2012-02-14).

68 Simon 1978 *Ariz LR* 673.

69 Book "Psychiatric Assessment for Rhinoplasty" 1971 *Arch Otolaryngol Head Neck Surg* 54; Reich 1975 *Plastic and Reconstructive Surgery* 11; Peterson & Topazian 1976 *J Oral Surg* 162; Simon 1978 *Ariz LR* 683.

terms of surgical and psychological outcomes.[70] The patient should therefore be questioned regarding his or her psychiatric history and current mental state.[71] The cosmetic surgeon should rather not proceed with the surgery if the patient is significantly depressed, psychotic or suffers from body dysmorphic disorder.[72] It is neither a cosmetic surgeon's duty to diagnose a patient with a psychiatric or personality disorder, nor does it fall within their scope of practice, however as part of the screening process the cosmetic surgeon should at least be aware of signs that a patient might be suffering from body dysmorphic disorder.[73] Although psychiatric treatments for the disorder can be effective, many patients who suffer from body dysmorphic disorder do not seek psychiatric help, instead they vehemently pursue a surgical solution for a psychological problem.[74] The cosmetic surgeon should inquire as to the amount of time spent each day worrying about the cosmetic defect, how much distress the perceived imperfection causes and whether concern over the imperfection has had any behavioural consequences such as social avoidance.[75] If the patient is preoccupied with the perceived imperfection to the extent that it causes significant distress or impairment in functioning, body dysmorphic disorder may be present.[76] Furthermore, if the flaw in the patient's appearance is far more insignificant than the patient perceives it to be, this might also be indicative of body dysmorphic disorder.[77]

5 A Physician's Legal Liability Concerning Patient Selection: Application of General Delictual Principles

A psychological or psychiatric injury can be described as any recognisable harmful infringement of the brain or nervous system of a person.[78] Psychological injury can be sustained in a variety of ways, including nervous shock, fright or other forms of mental suffering.[79]

70 Honigman *et al* 2004 *Plastic and Reconstructive Surgery* 1236.
71 Reich 1975 *Plastic and Reconstructive Surgery* 9; Simon 1978 *Ariz LR* 682.
72 Honigman *et al* 2004 *Plastic and Reconstructive Surgery* 1236.
73 Pruzinsky 1993 *Plastic Surgical Nursing* 109 ; Sarwer *et al* 1998 *Plastic and Reconstructive Surgery* 1644 ; Honigman *et al* 2004 *Plastic and Reconstructive Surgery* 1236; Pitts-Taylor 131.
74 Honigman *et al* 2004 *Plastic and Reconstructive Surgery* 1236; Newell *J Leg Med* 318; Newell 2011 *J Leg Med* 318.
75 Honigman *et al* 2004 *Plastic and Reconstructive Surgery* 1236 ; Crerand *et al* 2006 *Plastic and Reconstructive Surgery* 171; Newell 2011 *J Leg Med* 318.
76 *Ibid.*
77 *Ibid.*
78 *Bester v Commercial Union Versekeringsmaatskappy van SA Bpk* 1973 1 SA 769 (A) 775; *Barnard v Santam Bpk* 1999 1 SA 202 (HHA) 208-209; *Majiet v Santam Ltd* 1997 4 All SA 555 (K) 567; Neethling "Deliktuele Aanspreeklikheid Weens die Veroorsaking van Psigiese Letsels" 2000 *TSAR* 1.
79 *Minister of Justice v Hofmeyr* 1993 3 SA 131 (A) 145-6; Neethling 2000 *TSAR* 1.

Thus far courts have been primarily concerned with delictual liability due to the infliction of nervous shock, however there is no reason why psychological or psychiatric harm caused in ways other than emotional shock would not be actionable.[80] Delictual liability is primarily determined with reference to the requirements of wrongfulness, fault and legal causation. The brain or nervous system is as much a part of the physical body as any limb and any infringement of a person's physical-psychological integrity is regarded as *prima facie* wrongful and therefore actionable.[81] The cosmetic surgeon's duty to screen patients in order to prevent psychological harm is rooted in the fiduciary relationship that exists between a physician and patient.[82] This duty can be legally enforced by applying the general principles of negligence.[83] In order to establish negligence, the reasonable foreseeability and preventability of the psychological harm must be ascertained. The essence of negligence lies in the foreseeability of harm that may give rise to a duty to take reasonable steps to prevent the harm.[84] The foreseeability of emotional harm resulting from the performance of an elective surgery (particularly cosmetic surgery) without prior screening is well documented within the medical research.[85] This is partly due to the fact that, in the case of cosmetic surgery, underlying psychological considerations will always be present.[86]

When applying the reasonable foreseeability test it is impossible to formulate exact legal criteria for the determination of the reasonable foreseeability of the harm.[87] There are, however, a few broad and flexible guidelines that can be followed in order to determine the foreseeability of the harm. The general guideline is that the foreseeability of the harm is dependent upon the degree of probability of the manifestation of the harm.[88] If there was a very strong possibility of harm occurring, then it is highly likely that the resulting harm was

80 Neethling 2000 *TSAR* 1; *Barnard v Santam Bpk* 1999 1 SA 202 (HHA) 208-209.

81 Neethling 2000 *TSAR* 1 4; *Bester v Commercial Union Versekeringsmaatskappy van SA Bpk* 1973 1 SA 769 (A) 779; *Barnard v Santam Bpk* 209.

82 Nugent "Cosmetic Surgery on Patients with Body Dysmorphic Disorder: the Medical, Legal and Ethical Implications" http://works.bepress.com/cgi/viewcontent.cgi?article = 1004&context = Kristen_nugent (accessed 2012-02-14); Panfilov & Larkin 6.

83 Simon 1978 *Ariz LR* 673.

84 Neethling *et al* 125; *Kruger v Coetzee* 1966 2 SA 428 (A) 430.

85 Jefferson 1976 *J Nat Med Ass* 411-3; Simon 1978 *Ariz LR* 684.

86 Wright & Wright 1975 *Arch Otolarygol Head Neck Surg* 151; Simon 1978 *Ariz LR* 686; Mantese *et al* "Cosmetic Surgery and Informed Consent" 2006 *Mich Bar J* 27.

87 Van der Walt & Midgley *Principles of Delict* (2005) 177-8; *S v Bochris Investments* 1988 1 SA 861 (A); Visser *et al General Principles of Criminal Law Through the Cases* (1990) 541-63; *Butters v Cape Town Municipality* 1993 3 SA 521 (K); *Deysel v Karsten* 1994 1 SA 447 (A); *Cape Town Municipality v Butters* 1996 1 SA 473 (K).

88 *Ibid.*

reasonably foreseeable.[89] As far as the preventability of foreseeable harm is concerned, four factors must be taken into consideration. These factors relate to the probability of harm, the gravity of potential harm, the importance and object of the wrongdoer's conduct and the gravity of the burden of preventing the harm.[90] These factors help determine how a reasonable cosmetic surgeon would act under similar circumstances. When the burden of taking steps to prevent the harm is light and the probability and gravity of harm is great, a legal duty to take steps to prevent the harm exists.[91] The probability of harm may be relatively small, but the burden of making reasonable inquiries and referring doubtful cases to psychologists or psychiatrists to avoid serious harm is negligible.[92] Therefore, the cosmetic surgeon carries a legal duty to take these precautions and performing cosmetic surgery without paying any attention to the emotional fitness of the patient will constitute negligence.[93] In this regard Van der Walt and Midgley[94] state that:

> In general the magnitude of the risk must be balanced against the utility of the conduct and the difficulty, expense or other disadvantage of desisting from the conduct or taking a particular precaution. If the magnitude of the risks outweighs the utility of the conduct, the reasonable person would take measures to prevent the occurrence of harm; if the actor failed to take such measures he or she acted negligently. On the other hand, if the burden of eliminating a risk of harm outweighs the magnitude of the risk, the reasonable person would not take any steps to prevent the occurrence of foreseeable harm.

A cosmetic surgeon's decision to perform cosmetic surgery on a psychologically and emotionally unfit patient could even sometimes, in rare cases, be seen as a misdiagnosis.[95] For example, further cosmetic surgery will not cure the patient suffering from body dysmorphic disorder. Such a patient needs psychiatric help, not additional surgery. That being said, a misdiagnosis will not always constitute negligence. An erroneous diagnosis *per se* is not necessarily negligent, but the failure to adequately examine a patient in order to formulate an accurate diagnosis is indeed negligent.[96]

The patient plaintiff would also have to prove that a reasonably close nexus existed between the cosmetic surgeon's breach of duty and the ultimate harm suffered by the patient.[97] The cosmetic surgeon's breach of duty must therefore be the proximate cause of the harm suffered by

89 *Ibid.*
90 Simon 1978 *Ariz LR* 685; Neethling *et al* 139-40; Van der Walt & Midgley 179.
91 Simon 1978 *Ariz LR* 686.
92 Wright & Wright 1975 *Arch Otolarygol Head Neck Surg* 151; Jefferson 1976 *J Nat Med Ass* 412; Peterson & Topazian 1976 *J Oral Surg* 157; Simon 1978 *Ariz LR* 686.
93 Simon 1978 *Ariz LR* 686.
94 Van der Walt & Midgley 179.
95 Simon 1978 *Ariz LR* 673.
96 *Ibid.*
97 *Idem* 688.

the patient. In the case of psychological harm following surgery, the patient plaintiff would have to prove that had the surgeon refused surgery or warned of the imminent risks of psychological harm, the surgery would never have been performed and the resulting psychological harm would not have been incurred by the patient. The patient plaintiff would not have to prove that he or she had no pre-existing psychological problems. Cause includes substantial or material factors contributing to ultimate injury, therefore a plaintiff would only have to prove that the surgeon's conduct had been a precipitating and significant factor that triggered new problems or substantially worsened existing problems.[98]

As far as the nexus between the defendant's conduct and harm suffered by an abnormally psychologically vulnerable patient is concerned, the *talem qualem* rule also becomes relevant. This is a well-established legal principle and is commonly referred to as the egg-skull rule.[99] This rule is traditionally expressed in the maxim "the wrongdoer must take the victim as he finds him".[100] Egg-skull cases arise where the plaintiff, because of one or other physical, psychological or economic weakness, suffers a worse injury or loss as a result of the wrongdoer's conduct than would have been the case had the plaintiff not suffered from the particular weakness.[101] The egg-skull rule has its origin in the English case of *Dulieu v White and Sons*[102] where the court held that: "If a man is negligently run over or otherwise negligently injured in his body, it is no answer to the sufferer's claim for damages that he would have suffered less injury, or no injury at all, if he had not had an unusually thin skull or an unusually weak heart."[103] Legal scholars generally agree that in such a case the wrongdoer should be held liable for all the harm which may be ascribed to the existence of the weakness.[104] Just as a defendant who injures a plaintiff with an egg-skull is not entitled to use the abnormal vulnerability of the plaintiff's skull as a defence, it could be argued that a cosmetic surgeon cannot use the extraordinary psychological vulnerability or the egg-skull psyche of an unstable cosmetic surgery patient as a defence.[105]

There are several theories on how the liability of the wrongdoer should be justified, or which test for legal causation should be used to express

98 *Idem* 689.
99 Neethling *et al* 191.
100 *Hay or Bourhill v Young* 1943 AC 92 109-110.
101 Neethling *et al* 191.
102 *Dulieu v White and Sons* 1901 2 KB 669.
103 *Idem* 679.
104 Neethling *et al* 192.
105 Simon 1978 *Ariz LR* 690; Neethling 2000 *TSAR* 10. Van der Walt & Midgley 173 state that: "The doctrine of direct consequences has exerted its strongest influence on the question of liability for personal injuries. Once a defendant has been proved to have acted wrongfully and negligently, his or her responsibility embraces any harm flowing from a latent physical condition of the plaintiff, however unforeseeable or abnormal. This principle, inherent in the theory of direct consequences, is usually expressed by stating that the tortfeasor 'must take the victim as he finds him'."

liability in egg-skull cases.[106] Suffice it to say that the most acceptable approach to egg-skull cases is a flexible test for legal causation as illustrated in the case of *Smit v Abrahams*[107] where Farmal J investigated the rule and came to the conclusion that the fact that a plaintiff has a proverbial egg-skull is just one more fact that needs to be considered when applying all the other facts of the particular case in terms of the flexible test for causality. One must therefore determine whether, on the basis of reasonableness, fairness and justice, and in the light of all the circumstances of the case, the damage should be reasonably be imputed to the defendant.[108] Other existing tests for legal causation may also play a secondary role when determining legal causality in terms of the flexible approach.[109] In particular, the direct consequences test may be one of several factors taken into account in egg-skull cases.[110] In terms of the direct consequences test, a wrongdoer is liable for all direct consequences of his negligent conduct, irrespective of whether these consequences were reasonably foreseeable.[111] The fact that the psychological harm to an abnormally vulnerable cosmetic surgery patient was not reasonably foreseeable would in terms of this test not constitute a defence. The direct consequences test cannot be applied as a general test for causality.[112] It may simply, along with all the other tests for causality, play a subsidiary role. It may be a particularly relevant factor, but not the only factor when determining legal causality in egg-skull cases.

6 A Physician's Legal Liability Concerning Patient Selection: Assault Due to a Lack of Informed Consent

The cosmetic surgeon's duty to screen patients could also be enforced by applying the doctrine of informed consent. As a patient with a psychiatric disorder might possibly lack autonomy or the capacity to consent to treatment, he or she could be incapable of giving his or her informed consent to the treatment or surgery. Such a lack of informed consent due to a lack of capacity on the part of the patient could give rise to an action for battery against the surgeon. Capacity refers to competence; which is the functional ability to meet the demands of the decision-making situation by evaluating the potential consequences.[113] The fact that a patient's choice seems irrational, or does not accord with the cosmetic

106 For more information on the different theories of justification see Neethling *et al* 192-193.
107 *Smit v Abrahams* 1994 4 SA 1 (A).
108 Neethling *et al* 193.
109 Neethling *et al* 181; *Standard Chartered Bank of Canada v Nedperm Bank Ltd* 1994 4 SA 747 (A) 765.
110 Neethling *et al* 187; Van der Walt & Midgley 207.
111 Neethling *et al* 185; Van der Walt & Midgley 206.
112 Neethling *et al* 187 ; Boberg 442 ; Visser *et al* 112.
113 Carstens & Pearmain 879.

surgeon's view of what is in the patient's best interests *per se*, is not evidence that the patient lacks competence.[114] Furthermore, patients who have psychological problems or who are vulnerable to such problems are generally not mentally ill or incompetent to give consent.[115] Even mentally ill persons are generally able to consent to medical treatment as mental illness *per se* does not render a person unable to consent.[116] Mentally ill patients are only incapable of giving informed consent if their disorder prevents them from understanding what they are consenting to.[117] It is possible that severe psychological problems could render an individual incompetent to give consent to surgery.[118] This would be the case if the mental disorder prevents the patient from making definitive decisions, communicating his or her consent or accepting the need for medical intervention.[119]

7 A Physician's Legal Liability Concerning Patient Selection, Informed Consent and Body Dysmorphic Disorder: the Case of *Lynn G v Hugo*

An example can be found in the New York case of *Lynn G v Hugo*.[120] In this case, the plaintiff had visited the defendant, her cosmetic surgeon, nearly fifty times over a six year period to discuss various cosmetic surgery procedures. During that period, she underwent several cosmetic surgeries, including eyelid surgery, facial liposuction, eyebrow tattooing and wrinkle and skin growth removals. At some stage the plaintiff elected to undergo liposuction on her stomach. When the liposuction failed to produce the desired results, the plaintiff decided to undergo a full abdominoplasty to tighten her abdomen. Prior to the surgery, the defendant informed plaintiff of the risks associated with an abdominoplasty, including ugly scarring. The plaintiff acknowledged, in writing, her understanding of the risks and signed a consent form. Following her surgery, the plaintiff complained of an unsightly scar on

114 Par 9.2 HPCSA *Guidelines for Good Practice in the Health Care Professions, Seeking Patient's Informed Consent: The Ethical Considerations*.

115 Simon 1978 *Ariz LR* 691; Geffroy & Vernaglia "'But Are You Really Sure?': Requiring Psychiatric Proof of Patient's Informed Consent Prior to Elective Surgery" 2001 *Medicine and Health/Rhode Island* 142; Carstens & Pearmain 900; Cantor 2005 *Seminars in Cutaneous Medicine and Surgery* 159; Francis "Informed Consent in Body Dysmorphic Disorder" http://www.med scape.com/viewarticle/758800_4 (accessed 2013-02-20).

116 Carstens & Pearmain 900; Cantor 2005 *Seminars in Cutaneous Medicine and Surgery* 159; Francis "Informed Consent in Body Dysmorphic Disorder" http://www.medscape.com/viewarticle/758800_4 (accessed 2013-02-20).

117 Carstens & Pearmain 879; Francis "Informed Consent in Body Dysmorphic Disorder" http://www.med scape.com/viewarticle/758800_4 (accessed 2013-02-20).

118 Simon 1978 *Ariz LR* 691.

119 Carstens & Pearmain 901.

120 *Lynn G v Hugo* 96 NY2d 306.

her abdomen. The plaintiff then instituted legal action against the defendant, alleging a lack of informed consent and medical malpractice. Specifically, plaintiff claimed that she lacked capacity to consent to the procedures because she suffered from body dysmorphic disorder. The crux of her argument was that if a patient suffers from a mental disorder that is directly related to the procedure to which he or she is consenting, the validity of that patient's informed consent should be called into question.[121] The plaintiff claimed that her unusually high demand for surgical correction of minimal defects, together with the defendant's awareness of her use of antidepressant medication, should have alerted him to her condition, and that the defendant was negligent in not referring her to a psychiatrist before performing the surgeries. Furthermore, she asserted that body dysmorphic disorder had diminished her capacity to provide valid informed consent. The plaintiff neither contended that body dysmorphic disorder had rendered her incapable of concluding any contracts, nor that it had rendered her incapable of providing informed consent to surgery in general. The plaintiff simply claimed that body dysmorphic disorder had impacted her ability to properly evaluate risks and benefits of cosmetic surgery, as it had caused her to irrationally exaggerate her perceived physical imperfections. The court held that there was insufficient evidence to establish that the plaintiff actually suffered from body dysmorphic disorder. Consequently the court was silent on whether body dysmorphic disorder could potentially invalidate a patient's ability to provide informed consent to cosmetic surgery. Regardless, the case is still significant as the court's review of the adequacy of evidence to establish whether the plaintiff suffered from body dysmorphic disorder indicates that body dysmorphic disorder might potentially influence the validity of a cosmetic surgery patient's informed consent.[122] The court of appeal avoided making a pronouncement on the lower court's holding that cosmetic surgeons should be aware of established psychiatric conditions that affect body image and could impair a patient's ability properly to evaluate and consent to cosmetic surgery. The lower court's analysis of the surgeon's method to obtain informed consent was also sidestepped on appeal. The consequence of the lower court's opinion, never rejected at appellate level, is that when a patient suffers from compromised judgment, the determination whether or not to undergo surgery cannot exclusively be left with the patient.[123] Despite the fact that the court of appeal did not apply the doctrine of informed consent, the case is still significant as it is indicative of the fact that claims regarding informed consent by cosmetic surgery patients suffering from body dysmorphic disorder are beginning to emerge, that courts are starting to recognise that body dysmorphic disorder could potentially influence the validity of patient's consent and that cosmetic surgeons must take a more proactive

121 Mantese *et al* 2006 *Mich Bar J* 27; Pitts-Taylor 131; Newell 2011 *J Leg Med* 328.
122 *Idem* 325.
123 *Ibid.*

role in determining whether a patient should undergo cosmetic surgery when that patient's judgment is possibly impaired.[124]

8 A Physician's Legal Liability Concerning Patient Selection: Can a Physician Be Held Contractually Liable?

As far as a cosmetic surgeon's contractual liability for operating on a patient suffering from body dysmorphic disorder is concerned, courts are unlikely to hold a surgeon who operated in accordance with reasonable care and skill liable in contract for such a patient's poor physical or psychological outcome after surgery. That being said, if the court determines that the cosmetic surgeon, in operating on a patient ostensibly suffering from body dysmorphic disorder, did not treat the patient with a reasonable degree of professional skill and care and to a standard required by the professional and ethical rules of the profession, such a patient may have a claim based on a breach of contract. It is an implied or tacit term in healthcare contracts that the physician will treat the patient with a reasonable degree of professional skill and care and to a standard required by the professional and ethical rules of the profession.[125] Furthermore, as far as the formation and conclusion of a contract is concerned, informed consent is usually a precursor to the contract, but it may also form part of the terms of the contract itself.[126] It is generally an implied or tacit term in health care contracts that the patient's informed consent will be obtained with regard to treatment that is administered or surgery that is performed on the patient prior to the administration or performance of such treatment or surgery.[127] It is also worth noting that if the cosmetic surgeon had made a contractual promise of a certain surgical outcome to the patient, the issue of negligence and whether or not the patient had otherwise been competent to give informed consent is inconsequential with regards to the determination of the cosmetic surgeon's liability.[128] A representation which occurs during pre-contractual negotiations can be made part of the consensus between the parties and as such might become a term of the

124 *Ibid.*

125 Carstens & Pearmain 364. See Carstens & Pearmain 362 where it is said that there are a number of terms, including that a patient will be treated with a reasonable degree of professional care and skill, which may be inferred in a healthcare contract on the grounds of public policy, fairness and reasonableness. These terms are derived as much from the law of delict, constitutional law and administrative law as from the law of contract.

126 *Idem* 313.

127 Carstens & Pearmain 364; *Castell v De Greef* 1994 (4) SA 408 (C); *Broude v McIntosh* 1998 (3) SA 60 (SCA).

128 Nugent "Cosmetic Surgery on Patients With Body Dysmorphic Disorder: the Medical, Legal and Ethical Implications" http://works.bepress.com/cgi/viewcontent.cgi?article=1004&context=Kristen_nugent (accessed 2012-02-14).

ensuing contract.[129] This is relevant in the case of patients suffering from body dysmorphic disorder as such patients are more likely to interpret statements of opinion made by the cosmetic surgeon with some optimistic colouring, in their own minds thereby transforming such statements into promises and guarantees.[130]

A claim based on breach of contract might be particularly attractive in cosmetic surgery cases where the plaintiff has trouble recovering on other grounds. When courts are permitted to infer contractual guarantees from speculative oral statements made by the cosmetic surgeon during consultation, the cosmetic surgeon's scope of liability is broadened by bringing an action for breach of contract.[131] From the plaintiff's perspective, the possibility of pleading a case in contract is attractive since it renders his or her burden of proof significantly lighter as there is no need to prove negligence on the part of the cosmetic surgeon.[132] This suggests that the cosmetic surgeon's potential liability is increased under the contractual theory, while under negligence theory it is limited to actions where fault can be proven.

An action in contract is also particularly attractive if the cosmetic surgeon advertised or engaged in puffery during consultation in a manner that vulnerable individuals, such as those suffering from body dysmorphic disorder, might have interpreted as an express guarantee of satisfaction or of a certain aesthetic result.[133] Given the high percentage of individuals opting for cosmetic surgery who suffer from body dysmorphic disorder, telling a patient that he or she can surgically achieve the same appearance as a model or a celebrity, or even using computerised digital imaging software is not recommended as patients may accuse surgeons of warranting the surgeon's ability to achieve that particular result.[134] Not even the most skilled cosmetic surgeon can make a patient look identical to the prognostic virtual image, and the patient suffering from body dysmorphic disorder's distorted perspective will focus on any deviations from the aesthetic ideal he or she had hoped

129 Van der Merwe *et al Contract General Principles* (2007) 106.

130 Nugent "Cosmetic Surgery on Patients With Body Dysmorphic Disorder: the Medical, Legal and Ethical Implications" http://works.bepress.com/cgi/viewcontent.cgi?article = 1004&context = Kristen_nugent (accessed 2012-02-14).

131 Bonebrake "Contractual Liability in Medical Malpractice – *Sullivan v O'Connor*" 1974 *DePaul LR* 217.

132 *Ibid.*

133 Nugent "Cosmetic Surgery on Patients With Body Dysmorphic Disorder: the Medical, Legal and Ethical Implications" http://works.bepress.com/cgi/viewcontent.cgi?article = 1004&context = Kristen_nugent (accessed 2012-02-14).

134 Chávez *et al* "Legal Issues of Computer Imaging in Plastic Surgery: a Primer" 1997 *Plastic and Reconstructive Surgery* 1601; Koch *et al* "Advantages and Disadvantages of Computer Imaging in Cosmetic Surgery" 1998 *Dermatologic Surgery* 195; Nugent "Cosmetic Surgery on Patients With Body Dysmorphic Disorder: the Medical, Legal and Ethical Implications" http://works.bepress.com/cgi/viewcontent.cgi?article = 1004&context = Kristen_nugent (accessed 2012-02-14).

to achieve.[135] A disillusioned patient could easily grab on to what he or she understood as a guarantee of perfection and complete satisfaction to initiate litigation against the surgeon. Recovering with a contractual claim under these circumstances might require a strained interpretation of the surrounding facts and will probably only be used as a cause of action as a last resort when the battery due to a lack of informed consent and medical negligence charges have failed.[136] If, however, the court believes that the patient could have reasonably interpreted the cosmetic surgeon's statements during consultation, the surgeon's advertising materials or the surgeon's use of computer imaging as a promise to attain a certain result, the court may allow the patient to recover damages.

9 Conclusions and Recommendations

Cosmetic surgeons could only benefit from being acutely aware of the psychological undercurrents and possible psychiatric disorders such as body dysmorphic disorder, unrealistic expectations or heightened narcissism not remediable with surgery.[137] The success of a physician-patient relationship is based on good communication.[138] Cosmetic surgeons, more so than most other specialists, need to be careful listeners with clinical acumen that extends beyond the typical borders of medical illness.[139] Most problem patients can be handled effectively by the cosmetic surgeon by being gentle, but completely candid during the counselling interview.[140] If the cosmetic surgeon believes that there might be present or potential psychological disturbances, he or she should candidly confront the patient and recommend that the patient arranges a psychiatric consultation prior to surgery.[141] The cosmetic surgeon is under no duty to cure the patient, but he or she must at least refer the patient to a specialist for counselling before even considering the possibility of proceeding with the surgery.[142] Implementing a basic screening process will not only promote the patient's best interest; it will also protect the cosmetic surgeon against unnecessary and malicious litigation.

135 *Ibid.*
136 Nugent "Cosmetic Surgery on Patients With Body Dysmorphic Disorder: the Medical, Legal and Ethical Implications" http://works.bepress.com/cgi/viewcontent.cgi?article = 1004&context = Kristen_nugent (accessed 2012-02-14).
137 Atiyeh *et al* 2008 *J Aesthetic and Plastic Surgery* 831; Ericksen & Billick 2012 *Psychiatric Qly* 2.
138 Sykes 2009 *Curr Opin Otolaryngol Head Neck Surg* 321; Koch *et al* 1998 *Dermatologic Surgery* 1601.
139 Sykes 2009 *Curr Opin Otolaryngol Head Neck Surg* 321.
140 Wright & Wright 1975 *Arch Otolarygol Head Neck Surg* 151.
141 Book 1971 *Arch Otolaryngol Head Neck Surg* 54; Simon 1978 *Ariz LR* 683; Aronoff 97.
142 Simon 1978 *Ariz LR* 688; Pitts-Taylor 140; Sykes 2009 *Curr Opin Otolaryngol Head Neck Surg* 324.

Restricting freedom of speech or regulating gatherings?

Paul Hjul

OPSOMMING

Beperking van Vryheid van Spraak of Regulering van Byeenkomste?

Hierde bydrae is 'n kritise oorsig oor Suid Afrikaanse wetgewing met betrekking tot optogte en betogings naamlik die *Regulation of Gatherings Act 205 of 1993*. In hierdie kritise oorsig word openbare geweld en die gevolge daarvan bevraeteken, die waarde van openbare optogte in verband met die opbou van demokrasie bespreek en die funksie en rol van die howe in die bevordering van vryheid van uitdrukking en assosiasie word vasgestel. Die gevolgtrekking is dat die *Regulation of Gatherings Act 205 of 1993* gewysig word moet om vryheid van uitdrukking en assosiasie te beter te beskerm.

1 Introduction

The state transformation that occurred in South Africa in the 1990s was (and resulted in) a shift in the manner of thinking on many political and jurisprudential issues. Included in this transformation was a change in the legal position of political demonstrations which ceased being viewed as an assault on the state – and as such liable to suppression by the police force. This shift occurred within a backdrop of extensive public violence and intimidation which undeniably influenced the nature of the transformation of South African law on political demonstrations. Political thinking, however, is inherently fluid and ever changing; therefore legislation which is crafted within a particular jurisprudence frequently finds itself improperly implemented and grossly misunderstood in a context where a different polity has gained dominance.

An excellent example of progressive legislation and jurisprudence being out of touch with the implementation regime can be found in South African law on political gatherings and demonstrations, which are statutorily "regulated"[1] by the Regulation of Gatherings Act[2] (RGA). This Act was drafted on the recommendations of the Goldstone Commission (discussed later) during the transition period[3] and unfortunately all

1 The extent to which legislation, as opposed to human nature (or crowd psychology), can ever regulate gatherings is a concept which needs to be critically considered. For the purposes of this article however the presumption is that legislation can both legitimately and effectively regulate crowds if properly implemented.

2 205 of 1993.

3 For the purposes of this article I am assigning the term "transition period" to refer to the period between 1990-02-02 (the announcement by then President FW de Klerk of the unbanning of the ANC and release of Nelson
3 Mandela) until 1997-02-04 (when the majority of clauses in the 1996-Constitution came into effect).

evidence on the implementation of the Act suggests that the attitudes and understandings of municipal authorities (despite their importance under the Act) have not enjoyed the same changes as the legal framework – resulting in many gatherings being unlawfully obstructed by local government.[4]

Additionally, since the inception of the *South African Police Service Act*[5] in 1995 the police have undergone such turmoil and policy inconsistency from senior management that endemic institutionalised police corruption has been allowed to fester.[6] The greatest casualty of these various conflicts in South African political engagement,[7] and institutional schizophrenia,[8] is the Police Service and its members. The implications of impaired policing on the implementation of the RGA cannot be overstated particularly because of the constitutional responsibility held by the police for the maintenance of public order.[9]

Despite its title the RGA regulates more than just gatherings[10] and it prescribes various aspects of conduct regarding all political demonstrations as well as certain gatherings that are not politically inclined. To this end the RGA fundamentally operates as a limitation on the constitutional rights to peaceful assembly and expression and simply imposes certain presumptions about the behaviour of crowds. Further the RGA grants powers to the police in gatherings regardless of whether those gatherings occur within compliance of the RGA.[11]

---

4 The principal misunderstanding which municipalities appear to harbour is that they are "empowered" by the RGA to determine whether "applications" brought under the RGA should be given "permission".

5 Act 68 of 1995.

Faull and Newham *Protector or Predator?: Tackling Police Corruption in South Africa* (2011) 20 *et seq*

7 Nathi Mthethwa (the minister responsible for policing) has before the National Assembly made the curious statement "I have instituted a task team, led by the state law adviser, to investigate such allegations, because they are so serious as to suggest the meddling of policing functions in politics", the implications of a culture wherein political authority is subject to police meddling cannot be overstated. Business Day "Mdluli faces probe, shifted sideways" 2012-05-10 (online http://www.businessday.co.za/Articles/Content.aspx?id = 171370 (accessed 2012-08-04).

8 Burger "Institutional Schizophrenia and Police Militarisation" 2010-09-14 *ISS Today* (online http://issafrica.org/iss-today/institutional-schizophrenia-and-police-militarisation (accessed 20131025).

9 S 205(3) SA Constitution provides: "The objects of the police service are to prevent, combat and investigate crime, to maintain public order, to protect and secure the inhabitants of the Republic and their property, and to uphold and enforce the law."

10 Defined in the RGA to cover most forms of assembly and procession in public roads and public places in open air where more than 15 people are present.

11 S 9(1) RGA provides: "If a gathering is to take place, whether or not in compliance with the provisions of this Act a member of the police – ... (f) shall take such steps, including negotiations with the relevant persons, as

11 are in the circumstances reasonable and appropriate to protect persons and property, whether or not they are participating in the gathering or demonstration."

This article will give attention to sections 6, 7 and 8 of the RGA and questions the value of the provisions (or absence of provisions) relating to: (a) demonstrations outside of court buildings, Parliament and the Union Buildings; (b) the prohibition of the wearing of masks; (c) the prohibition of the blockading of access to any building or premises; (d) the utilisation of violent behaviour in demonstrations; and (e) the continued unlawfulness of contravention of other law during a gathering or demonstration.

For this article an evaluation of section 11 is omitted in its entirety because it, together with the potential alleged "chill effect", is a subject warranting separate attention.[12] This article is focused on organised intended peaceful gatherings as opposed to spontaneous (or allegedly spontaneous) gatherings and riotous violent assemblies.[13]

It is the absence of provisions permitting extra-ordinary demonstrations on application to the appropriate court which will form the focus of this article's critique. The argument that I present and sustain is that rather than attempting to protect the Courts from demonstrators, South Africa's constitutional dispensation requires the creation of a framework in which courts are able to be – and are seen to be – the guardians of free expression. This is something which the RGA fails to do. I therefore propose an amendment to the RGA to give the district magistrates courts power to authorise proscribed conduct in order to promote free expression.

2 Overview of the Act and its Underlying Assumptions

The RGA is divided into several sections each of which addresses a different aspect of the regulatory framework: sections 1 and 2 provide definitions and related matters; sections 3, 4 and 5 regulate gatherings, setting out conditions of notification for gatherings, the consultative process and the prohibition of gatherings; section 6 creates the structures by which the decisions taken under sections 4 and 5 may be set aside by a magistrate utilising an urgent procedure; section 7 proscribes both gatherings and demonstrations within 100 meters of court-buildings, Parliament and the Union Buildings other than with the written

12 The Constitutional Court ruled in *South African Transport and Allied Workers Union and Another v Garvas* [2012] ZACC 13 that this section of the RGA is constitutional. The Court in essence followed the Supreme Court of Appeal in *South African Transport & Allied Workers Union v Garvis* [2011] ZASCA 152. It is my opinion that the challenge by SATAWU was misconceived and a constitutional challenge would have been better placed at the definition of "riot damage" in the RGA. Sadly the effect of the decision of the Constitutional Court and surrounding discussions and literature may have a "chill effect" on gatherings and demonstrations.

13 To which end the incidents of rioting by the SANDF in 2009 are specifically excluded.

permission of the relevant magistrate (or director-general in respect of the Union Buildings); section 8 prescribes conduct for both gatherings and demonstrations; section 9 provides additional powers to the Police (the Act specifically does not remove any powers from the Police); section 10 makes provisions for the promulgation of regulations; section 11 imposes liability for riot damage on organisers of demonstrations and gatherings; section 12 creates various offences resulting from contravening the RGA; section 13 directs the interpretation of the RGA; and the remaining sections provide for miscellaneous matters.

As indicated in the introduction this article will place its focus on sections 7 and 8 and to a lesser extent sections 6, 9, 12 and 13.

2 1 Focused Sections

In addition to planned demonstrations, section 7 of the RGA, as written, effectively proscribes groupings of more than 15 people outside of court buildings regardless of the purpose for which they are gathering and proscribes all political demonstrations outside of Courts without prior written permission. This is because the definition of gathering incorporates "any assembly, concourse or procession". Therefore should the media form a gathering outside of the Johannesburg, Pretoria or Pietermaritzburg high court in which more than 15 persons are acting in concert they fall foul of the provisions of the RGA. It is unlikely that any attempted prosecution of this contravention will ever take place and, should adherence to the letter of the law actually occur, I suspect that the courts will be even more overwhelmed with work than they presently are. Authorising[14] the magistrate's court to *suo moto* grant permission, provide exemptions, or to designate a space for demonstrations would eliminate this potential for difficulty.[15] The amendments proposed in this article eliminate the need entirely.

Sections 8 and 9 of the RGA apply to all functions (that is gatherings and demonstrations) that fall within the ambit of the RGA and it is therefore these sections that warrant attention to discussions on political protests generally. The first aspect of section 8 that is apparent is the obligatory rather than recommendatory nature of its terms – in actual fact the RGA at no point establishes a system of devising recommendations for organisers such as to remain within the ambit of the RGA – meaning that rather than aiding in the regulation of demonstration it proscribes specific "undesirable" conduct. When section 8 is read together with section 12 the purpose is clear: the RGA

14 As this authorisation is essentially an administrative issue I believe it can and should be provided for by amendment to the (R 361 of 1993-03-11); the provisions of reg 13, relating to media access are an apt starting point for the proposed amendment.

15 The issue is largely a matter *de minimis*, however, problems will occur if the state prosecutes grass-roots organisations protesting outside a magistrate's office while allowing a political party to conduct a press conference in the same space.

renders it a criminal offence to conduct gatherings and demonstrations in a particular manner.

2 2 Missing Provisions and Structural Bias

The RGA holds a bias towards the organisation of mass gatherings in the stead of small demonstrations. This is attested to by the joint effect of the provisions of sections 6 and 9. Whereas decisions taken by the responsible officer in connection with restricting or prohibiting gatherings are subject to judicial oversight, no process of automatic intervention by the courts apply where a police officer utilises the powers afforded under section 9 of the RGA. This places excessive responsibility, and power, in the hands of police officers with varied levels of training and specialisation. Therefore, a legitimately organised demonstration may find itself dispersed as a result of over-cautious policing even if the organisers act in good faith simply because the police have not had the opportunity to engage with the organisers in a meeting called by the responsible officer. This is because demonstrators are not afforded an opportunity to acquire any form of approval,[16] and it may well occur that a police officer may believe, in good faith, the demonstration to be illegal – a situation aggravated by repeated incorrect and/or problematic media reports and quotes from municipal and political party spokespersons.[17] Further, as will be argued in the duration of this article, the "disruptive" requirement for a successful demonstration coupled with civic complacency makes the mass march of unthinking participants a far less challenging initiative than imagining and executing a well thought out and expressive demonstration.

2 3 Function Performed by the RGA

In essence the RGA performs a triple regulatory function. It regulates (a) all public open air gatherings; (b) certain activities containing less than 15 people where those activities fall broadly within the scope of protest action; and (c) the presence of persons within 100 meters of court-buildings, Parliament and the Union Building. Fundamentally the act merely upholds freedom of assembly and does little with respect to Freedom of Expression. It is my submission that it is really only the first of these functions that is done in a satisfactory manner. Furthermore, this failure results in a climate in which the mass march is the sole means of peaceful non electoral political expression holding dominance in South Africa – and that this contributes to a general disenfranchisement and ultimate disengagement at a grass-roots level with peaceful protest action.

16 Glasner *Mbeki and after: reflections on the legacy of Thabo Mbeki* (2010) 114, 122. While the RGA operates on a jurisprudence of giving notice rather than seeking approval, the general question of "who approved this", "who gave you permission" permeates human thinking particularly amongst police officials.

17 Glasner 120.

3 The Focus of Public Violence

Violence is seldom, if ever, spontaneous, but arises from a conviction that fundamental rights are denied.[18]

3 1 Delegitimising the Legitimate Rebellion of the Poor

Service delivery protests, which are often accompanied by violence – and generally contravene the RGA – are an ingrained feature of grass-roots resistance against public corruption, ineptitude and a disregard for the poor by those in power. Unfortunately all accounts suggest that the South African government will only take cognisance of certain communities when those communities resort to violent protest in a phenomenon described as a "rebellion of the poor".[19] However by resorting to violent protests the communities expose themselves to further violent suppression by the police and are further maligned by political authorities as "violent" and "criminal".

The fact is that protesters are able to resort to violent means of ensuring success in garnering support or understanding for their particular cause. Therefore in addition to the political and social implications legal literature should not ignore the interactions between defiance and compliance in moulding the legal system.[20] A preoccupation with preventing public violence and preserving "the peace" as part of law and order readily overtakes the constitutional imperative of establishing a framework in which free expression and exercise of political rights can take place. It simply does not suffice to uncritically regard transgressors of a non-violence ethic as delinquent or to assume that the violence is an inevitable consequence of "class warfare" – whatever that may mean. The RGA falls short in that, while it prevents intimidative demonstrations, it does not encourage or support innovative demonstrations. In the South African context our laws relating to public gatherings are rooted in escaping a legacy of public violence, but it is the same South African context that demands an innovative approach to demonstrations as an instrument of democracy. In spite of Devenish's proclamation that the maxim of *salus republica suprema lex* is dead [21] the nature of the state results in an ingrained aversion to violence by any non-state actor, a fact that is manipulated by political authorities while trampling the dignity of "violent" opponents.

---

18 Herman "Injunctive Control of Disruptive Student Demonstrations" 1970 *Virginia LR* 215, 218 quoting the Commission on Industrial Relations (1916).

19 Alexander "Rebellion of the poor: South Africa's service delivery protests – a preliminary analysis" 2010 *Review of African Political Economy* 25.

20 While there are comparatively few publications in journals dedicated to law, the American student protests during the civil rights movement of the United States have received coverage: Herman 1970 *Virginia LR* 215; Anon "Regulation of Demonstrations" 1967 *Harvard Law Review* 1773.

21 Devenish, "The Demise of *salus reipublicae suprema lex* in South Africa: Emergency Rule in terms of the 1996 Constitution" 1998 *CILSA* 142.

Ultimately the political authorities are placed in the most unjust position by having at their disposal the capacity to infringe on the fundamental *dignas* and *fama* of those who oppose them. A trade off, between violent confrontation accompanied with "unconstitutional", "violent" and "irresponsible" behaviour and accepting the terrible status quo, is presented to grass-roots movements without any accountability attaching to the failed political authorities.[22] Coupled with this is the fact that the divisions between the interests of the State and the interests of the ruling party, or a faction thereof, are frequently blurred.

3 2 Strengthening State Instruments to Keep the Peace

The RGA has not replaced various other pieces of legislation or the common law regarding the preserving of the peace. Fundamentally magistrates and law enforcement remain charged with keeping the peace and our Criminal Procedure Act[23] is peppered with discourse arising from this charge. As is the fairly simple Justices of the Peace and Commissioners of Oaths Act[24] which is a cornerstone in our legal system which remains un-amended in our constitutional discourse. Section 13(1)[25] also makes it clear that the RGA does nothing to interfere with the provisions of existing law regarding the prevention of public violence – or anything other than traffic laws and municipal ordinances controlling public space[26] for that matter.

Section 9(3) enhances police powers to use the State's monopoly of force even though no form of state of emergency exists. Thus, whilst the rights to assemble, march, and protest may be upheld in the RGA, it is the right of the police to use force to control those who exercise these rights that is really achieved.

3 3 Goldstone Commission

Earlier in the article it was said that the genesis of the RGA was found in the Goldstone Commission. This Commission was formed in 1991 as part of the process of establishing a climate in which a constitutional transition was possible. The Commission was not a Commission of Inquiry convened by the President in terms the Commissions of Inquiry

22 A comprehensive account of oppression of dissent during the Mbeki presidency is provided by Duncan "Thabo Mbeki and Dissent" in *Mbeki and after: reflections on the legacy of Thabo Mbeki* (ed Glasner) (2010) 105-127.

23 51 of 1977.

24 16 of 1963.

25 In some respects my argument relies quite heavily on s 13(1)(c) RGA. I have not, however, specifically evaluated it due to the fact that its importance is seen in the absence of other provisions rather than by its presence.

26 No case law appears to have been developed on the interaction of the RGA on other laws not specifically mentioned in s 13(1). I submit that if a gathering is planned and in the notice specific mention is made of actions which infringe local by-laws or policy pertaining to the usage of public space the operation of the RGA would render the by-law inoperative. However the proposal in this article would address the issue definitively.

Act [27]– such as the Hefer and King Commissions – but rather was mandated by the unrepealed[28] Prevention of Public Violence and Intimidation Act[29] which created in statute a Commission intended for more than the three year life span of the Goldstone Commission. The Commission ceased to exist due to the fact that no Judge was appointed to replace Judge Goldstone at the end of the three year term. I firmly believe that this failure to retain the Commission played a significantly detrimental role in the resurfacing of incidents of public violence including the security guard labour action violence, the xenophobic attacks in 2008, and the SANDF rioting in 2009.

The work of the Commission was extensive and the amount of material it covered is remarkable. However it must be borne in mind that the Commission was established to investigate public violence in a particular context – which has changed – and to recommend on methods of preventing and curtailing public violence. A bias of focusing on the prevention of public violence and preserving the peace is thus evident in the multitude of reports produced by the Commission.[30] The RGA, and all legislation relating to curtailing public violence, when it breaks out, is entirely unrelated to the root causes of public violence, which principally is inequitable social conditions and a political culture which ignores those in desperate need of political space.

3 4 Incomplete Transition

The rise of public violence in South Africa ten years after the conclusion of the Goldstone Commission is testimony to the fact that the transition from an apartheid era of violence into a human dignity based society is fundamentally incomplete. The fact that the current government in so many respect mimics and has returned to the policies of the apartheid era government bears further witness to the incomplete transition. Further it is indubitable that a resurgence of public violence is related to, and aggravates, an erosion of the Rule of Law.

The importance of providing a framework in which public engagement in non-harmful demonstrations occurs is not limited to the importance such engagement plays in democracy but extends into the realm of serving as a valve against more harmful forms of demonstrative behaviour (culminating in riots).[31]

---

27 Act 8 of 1947.

28 This act has fallen into disuse as a result of a failure by the Office of the President to appoint a successor to Goldstone J at the termination of the three year lifespan of office. I submit that it may be argued that the President has in fact failed in fulfilling a constitutional obligation to give the legislation effect but in reality the decision was taken on broad consensus at the time and it may equally be argued that the decision of *South African Association of Personal Injury Lawyers* v *Heath and Others* 2001 1 SA 883 (CC) requires a statutory reconsideration of the empowering act.

29 Act 139 of 1991.

30 The documents of the Commission are preserved by the Human Rights Institute of South Africa.

4 Value of Demonstrations in the Political Process – Consensus Building

The *Harvard Law Review,* in analysing the regulation of demonstrations in the United States as at 1967, poignantly stated that:

> Often a demonstration has significant publicity advantages over more conventional media of expression since it can attract extensive news coverage and widespread public interest; and for persons unpopular or unknown to the general public, or without financial resources, a demonstration may be the only effective means to publicise a message or reach a desired audience.[32]

Now whilst the term "demonstration" in the American usage is "gathering" in the RGA[33] the message is applicable, namely that unconventional or alternative expression channels are an invaluable tool in making democracy democratic. In the South African context the abundance of poverty is of specific significance – as evidenced by the growth in grass-roots service delivery protest actions. While the South African legislative process creates an array of public involvement opportunities many of these opportunities provide little help to individuals alienated from the government. It is for this reason that the strength of the RGA in its present form lies in its general holding of a right to participate in gatherings.[34]

4 1 The Peculiarity of those United States of America

The United States of America is a federal state employing as its basis for state law a notion of common law imported and modified (frequently by statute) from England.[35] The United States Supreme Court is charged with ensuring that all states uphold the Constitution of the United States. The Bill of Rights (the first ten amendments to the US Constitution) has effectively established a process by which the various idiosyncratic laws and practices of individual states in the United States are evaluated and

31 The RGA defines "riot damage" in a manner which I submit loses sight of the essence of what rioting is and has the unfortunate effect of failing to differentiate between an unruly crowd and a riotous assembly. Consider *South African Transport and Allied Workers Union and Another v Garvas* [2012] ZACC 13 particularly paragraph 95.

32 1967 *HLR* 1773.

33 In both American and British usage a demonstration containing few people is referred to as a "picket" regardless of whether conducted in a labour context or not.

34 This in my view is the saving grace of the RGA if not its implementation.

35 Louisiana is generally an exception to this rule. However Louisiana has adopted the English concept of breach of peace. Thus for the purposes of this article the rule can be held.

often struck for falling short of the provisions of that Constitution.[36] Thus, a multitude of legislative authorities enact laws regulating gatherings – usually proscribing political demonstration through statutes for trespass[37] or keeping of the peace.

The South African Constitution is different to the US Constitution in many respects however, key concepts can be extracted from US Jurisprudence. *The Harvard Law Review* analysis provides a starting point on extensive coverage of different situations and legislation – in this regard the federal nature of the United States is a help in furnishing comparative literature, as it provides a broad range of challenges to the US Courts. Furthermore the worry that, due to the resource imbalance in South Africa, a minority[38] is able to exert a disproportionate amount of political power is not unique to the South African context and is an invariable feature of any constitutional democracy.

Therefore the sentiments expressed in the *Harvard Law Review* that:

> Whenever the demonstrators are complaining of a bona fide wrong, society interests will be advanced if their grievance is brought to the public attention and relief granted; and the danger that demonstrating minorities may possess disproportionate political power would seem to be outweighed by the desirability of keeping the majority informed of minority complaints.[39]

are of great value to our fledgling democracy.

The present legal and political situation in South Africa does not prevent a militant mobile and unimaginative minority from exerting the same undue influence – on an art gallery for example. A gathering of 10,000 South African's still represent a fraction of a per cent of the population.

5 Should Demonstrations be Disruptive?

The legitimacy of regulating gatherings and demonstrations such as to prevent violence does not however automatically justify regulating gatherings and demonstrations such that they are not disruptive – what is the purpose of a demonstration other than to, in one manner or another, disrupt the intended audience such that they engage with the demonstration and its message? This is not to suggest that unduly disruptive behaviour should be permitted but rather that an evaluation of the purpose of the demonstration and the extent of disruption should be evaluated. More significantly, there is a pressing need to formulate a nuanced and critical understanding of what constitutes an actual threat to peace and safety in substitution of a superficial "peaceful gathering"

36 Within US jurisprudence the courts – particularly the federal courts "enjoin" state authorities against giving effect to unconstitutional law.
37 *Adderley v Florida* 385 US 39.
38 In the South African context the "white minority".
39 1967 *HLR* 1773.

approach. Is the expression of hate speech and the advocacy of causing harm against individuals or a section of society by a politically influential person not more riotous than the blockading of a road by marginalised people? The politically connected person will be afforded full protection by the state and may be held accountable by protracted legal proceedings whereas the grass-roots movement which blocks a road may be afforded the use of rubber bullets and water cannons on their persons.

Furthermore if we accept the view that:

> Functionally, disruptive demonstrations can be viewed as part of the process of social bargaining – a means of influence by groups who lack more positive methods of persuasion.[40]

It stands that the extent to which positive remedies are unavailable or felt to be unavailable will directly influence the extent to which violence will be resorted to. Thus enabling demonstrators to have the ability to feel that they are heard is valuable.

It is fortunate that the RGA does not give cause to prohibit any particular gathering or demonstration on the grounds of mere disruption but rather requires "serious disruption", furthermore the possibility of damage to some property is similarly not a definitive ground for prohibition.[41] This has not however prevented political authorities from seeking to suppress demonstrators and movements which challenge the cartel relationship currently enjoyed by political actors in South Africa.

6 The Value of (Injunctions) Interdicts and Judicial Sanction

Todd[42] places focus on the fact that the RGA provides rights to private persons who are negatively affected by gatherings – particularly riot damage – however, these rights are, in essence, not highly accessible. At the inception of the RGA a person (other than the authorised member of the police) was required to approach the high court (in terms of section 6(5) of the RGA) in order to have a condition imposed on a gathering.[43] The shift in the powers and procedures of the magistrates' courts – including the replacement of the old application rules (rule 55) with

40 Herman 1970 *Virginia LR* 215 216.
41 S 5(1), S 4(b) RGA.
42 Todd "Reading the Riot Act" 1994 *Juta Business LR* 93.
43 The RGA merely requires that an application must be brought to an "appropriate court". However read together with the prescription of making the application as "an urgent application in accordance with the Uniform Rules of the several Provincial and Local Divisions of the Supreme Court of South Africa" as well as the performance nature of such an order, the jurisdiction of the magistrates court would be difficult to establish and the clear purpose of "appropriate court" is to empower the labour courts.

provisions which mirror[44] the Uniform Rules applicable in the high court suggests that the magistrates' courts have definite jurisdiction to grant applications imposing conditions on gatherings. In reality however ordinary people suffer from the Ritz nature of the courts – particularly the High Court.[45]

This provision ties in to the American experience with disruptive gatherings. Certain demonstrations resulted in various resourced institutions approaching the courts for an injunction. Such injunctions were sought even where the conduct they were aimed at preventing was already illegal.[46] The reason for such applications lay in the various advantages of such an approach, most significantly the fact that "the actions of the police are cloaked with the court's prestige"[47] and the inadequacies (particularly in enforcement) of the existing laws and regulations proscribing the conduct. Ultimately if a gathering is placed under interdict and proceeds nevertheless, it being viewed as riotous (consequent upon being in contempt of court) should give affected persons greater rights.[48]

It is also possible to argue that the provisions of the RGA empowering the responsible officer to prohibit a gathering amount to bestowing the quasi-judicial power of injuncting an otherwise legal gathering.[49] However the grounds for prohibition available to the responsible officer are significantly more restrained than the available conditions for an interdict.

However, the cloak of judicial prestige is one that must systematically and periodically be lengthened. In the present South African context assaults on the judiciary have reached epidemic proportions. With regard to regulating demonstrations, the courts should be afforded an opportunity to be seen to be the guardians of freedom of expression. And in the same manner as they should be able to interdict gatherings and demonstrations they should be able to interdict and aide in preventing

44 To a large degree, again this is a subject warranting discussion. A particular possible problem arises in the extent to which magistrates' courts are able to operate outside of the rules and practice directive in contrast the High Court.

45 "Justice is open to all – like the Ritz hotel" Sir James Matthew (the exact origin of the amorphism is a subject of debate but the scope of attribution to the judge includes the *Collins English Dictionary* online http://www.collinsdictionary.com/dictionary/english/justice (accessed 2013-10-25)).

46 Herman 1970 *Virginia LR* 215 222.

47 *Idem* 223.

48 Consider *South African Transport and Allied Workers Union and Another v Garvas* [2012] ZACC 13 par 95 this may become illusionary, with the court arguably having upheld the constitutionality of the definition of riotous.

49 Consider s 5(1) RGA which specifically requires credible information brought under oath to exercise the power in essence the bringer of such credible information is an applicant for injunction. Somewhat perversely an applicant approaching the "appropriate court" does not have to satisfy the same standard as set out in s 5(1) "that the Police and traffic officers in question will not be able to contain this threat".

disruptions and threats against disruptions. It is perhaps a pity that none of the grass-roots movements who have found themselves in the firing range have visibly and successfully utilised judicial proceedings to secure an interdict against aggression from municipal law enforcement, private security firms or partisan groupings who introduce violence into gatherings.

Finally, problems for the cloak of judicial prestige arise from the fact that the RGA performs the function of ensuring the designation of a person as responsible for a particular gathering and thrusting onto them a particular liability framework. In this way a court is able to hold individuals accountable for civil wrongs – even with riot damage – but ignores the conduct of state actors.[50]

7 Transformation of Policing of Gatherings

The continuous institutional transformation, retransformation and reform imposed on the SAPS, together with the high turnover of national commissioners, makes evaluating how the SAPS police gatherings a complex exercise which warrants separate extensive coverage. However the fundamental question to consider is whether the SAPS have the capacity and nature to enforce the RGA and what policies and approaches towards the issues relating to public protest best align with a police service in a constitutional republic.

Commentators cautioned about the weakening of the capacity of the SAPS in public order policing as a result of various "costly" restructurings[51] during the reign of Jackie Selebi as national commissioner, and for this reason it was of even greater significance that the nature of demonstrations that take place be light on policing resources. With the rebellion of the poor starting in 2004 the weakening of SAPS capacity should have been accompanied by an increase in avenues and fora to promote non-violent actual engagement, this did not happen. 2008 witnessed the horrendous xenophobia related violence which required the deployment of the SANDF in order to maintain public order.[52]

In 2009 a politically mandated shift brought a return to militarism within the SAPS – which is particularly marked by the reintroduction of

50 COSATU in argument for *South African Transport and Allied Workers Union and Another v Garvas* [2012] ZACC 13 brought attention to the deficiency or perceived deficiency of state organs, particularly the SAPS, in preventing damages from occurring.

51 Oma *SAPS Costly Restructuring: A review of public order policing capacity* (2007).

52 On 2008-05-21 then President Thabo Mbeki authorised the deployment of SANDF to assist the police in quelling the violence. A comprehensive report on the violence was prepared by the Human Sciences Research Council *Citizenship, Violence and Xenophobia in South Africa* (2008) http://www.hsrc.ac.za/Document-2807.phtml (accessed 2012-08-04).

military ranks to the police service and the increased references to a police force. Further in 2010 the SAPS successfully discharged a massive exercise of public order and control in the form of the 2010 FIFA World Cup. These developments have however made no contribution towards aligning the SAPS with the protection of protestors or the promotion of free expression. Ultimately the prestige of the SAPS has been severely dented by successive scandals involving maladministration, brutality and corruption.

Establishing a framework in which free expression and political activity is promoted is not only possible but desirable for the prestige of the SAPS and its staff - the enhancement of the professional image of police officers (including metropolitan police and traffic police) as more than mere pedantic enforces of petty statutes is important. If the SAPS is to recover from the current taint that has been thrust upon it by the political climate of South Africa this should be the primary objective of senior management within the SAPS.[53]

8 Transformation of Local Government and the Problem of Municipalities

The Makana Municipality has structured an "application" form such as to approve gatherings; the Kouga Muncipality insist that a rather stringent 'Memorandum of Agreement" be entered into with the municipality; and the Mangaung Metropolitan Municipality require organisers to sign an indemnity for the municipality when seeking to hold a gathering. The attitude that the local political authorities determine who may protest appears to be the case across municipalities[54] and largely accounts for why certain grass-roots movements, on being wrongfully suppressed, resort to violent protest. Urgent intervention to capacitate municipalities to perform their mediatory function under the RGA is desperately required. Alternatively the RGA should be removed from the purview of municipalities entirely.[55]

53 Commissioner Nhlanhla Mkhwanazi who acted as National Commissioner during the suspension of Bheki Cele and before the appointment of Commissioner Riah Phiyega is the divisional commissioner responsible for public order policing. The commendable service of Commissioner Mkhwanazi gives some hope that with the support of the new National Commissioner will be able to establish a truly effective public order policing culture free of political interference.

54 The selection of these three municipalities simply emerges because each of the three represents the first instance of an approach to the RGA which I have encountered. Comprehensive research on municipal implementation of the RGA is a long overdue project of massive scope; see also Glasner 125.

55 For example in the United Kingdom notification must be given to the applicable police force and it may be questioned whether notification should be a requirement at all, for the purposes of this article I assume notification in itself not to be an infringement on free expression.

9 Improving the Provisions Pertaining to Demonstrations Outside Court Buildings

The provisions of Section 7 are not exceptionally onerous – although they have been made more onerous by the change in the application procedure applicable in the magistrates' courts[56] – but they do create problems: firstly the permission sought to demonstrate at a court-building is granted by the same court;[57] secondly the application process makes it difficult for a lay person to organise a demonstration or gathering (it must be added that there is no reason to assume assistance from court officials); thirdly the provisions effectively restrict access to the courts by the ordinary public wishing to express their views. Furthermore the courts (particularly superior courts) and their presiding officers should within the scope of regulating and upholding the integrity of their judicial process have the power to interdict conduct that undermines the judicial process, however, the holding of a demonstration outside of the physical premises of the court poses such a negligible threat to the judicial process and on the contrary promotes public engagement with the courts. Further a court can take judicial notice of the presence of demonstrators attending for a particular hearing without allowing such demonstrators to influence their decision. South African judges who have had the misfortune of having to preside at a high profile case have succeeded in balancing the interests of democratic engagement and the rule of law. I therefore submit that the present restriction system (rather than the provisions in the RGA) is unsatisfactory and is in fact almost unenforceable – many demonstrations and gatherings in contravention of the RGA do in fact take place.

10 Proposed Amendments

My proposed amendment are relatively simple: (a) insert the restriction pertaining to court buildings into Section 8 (as 8(11)); and (b) compose a new Section 7 in which an application to the district magistrate's court may be made such as to permit an allowance against the proscriptions in section 8 and/or the suspension of any particular "petty" laws or regulations in a particular public space provided that the applicant return the public space to the condition in which it was prior to the demonstration after its conclusion, and further provide that an official at the court must be designated to provide assistance in framing an application.

---

56 Rule 55 has been reworked in its entirety by Notice 740 GG 33487 of 2010-08-23 to resemble the motion court rules of the high court. The subject of the extent to which the change in rules has unintended consequences is warrants its own discussion.

57 As a technical matter I submit that the permission is not actually granted by the court but rather by the district magistrate in a quasi-judicial capacity.

This would mean that a magistrate's approval would be required for all demonstrations or gatherings in which demonstrators cover their faces or wear certain uniforms or re-enact scenes that involve the utterance of words of hatred or which take place within the confines of 100 meters of a court-building.

An additional consideration that may be made is to empower magistrates to authorise the service of a petition on some person or representative of an organisation. Presently demonstrators and gatherers have no legal right to having their petition received by any person and there is no method of compelling an official to take any cognisance of demonstrator's demands. While it could be highly problematic to compel a private person (or organisation for that matter) to accept a petition, it seems entirely in accordance with our constitutional democracy to empower the courts to enforce acceptance of petitions by officers of state on behalf of some or other organ of state or government. Such enforcement can take a formalised and ceremonious form – for example the relevant high ranking public official may be required to grant a short (less than 5 minutes) audience to a delegation of the gathering or demonstration of no more than 2 persons accompanied by a Police Officer or the Sheriff, in which the petition is formally served on the official.

11 Effect of Proposed Amendments

The proposed amendment establishes a procedure by which demonstrations can be organised in a manner that would otherwise be in contravention of some or another petty[58] law after being sanctioned by the Courts in advance on application. Therefore organisers of a demonstration would be able to apply to the Court to sanction a demonstration in which the intended message is of some importance (to the applicants in the very least) but which causes some form of disruption or contravention.

The Courts in determining whether to grant an application for an extra-ordinary demonstration will have to perform a balancing exercise. I believe that South African jurisprudence is sufficiently equipped to judiciously perform this task and the abundance of international experience should not be underestimated. In particular whilst jurisprudence on the First Amendment to the United States Constitution cannot be uncritically transposed to the South African context, the complexity of the United States experience, and the fact that the process of balancing freedom of expression and political activity against the keeping of the peace by the State, means that there is ample relevant lessons to draw. In preparing this article I have drawn heavily from the

58 This is not to suggest that the ordinary operation of such a law is unimportant but rather that the law can be suspended without causing serious harm to the functioning of the state. Most petty laws will be of the nature of municipal by-laws and regulations.

Regulation of Demonstrations a note appearing in *The Harvard Law Review* in 1967; it is clear therefore that more than 40 years of American experience is available. Furthermore there exists an abundance of American legal literature – rather than simply tomes of social science material.[59]

Furthermore the courts can not sanction any unilateral conduct which deprives people of their rights. Thus vandalism of property cannot possibly be permitted – whereas the covering of a particular wall or object in a manner that makes it appear vandalised would be.

The decision to locate this power in the magistrate's courts may create problems and complications, however if consideration is given to the various empowerments already bestowed onto magistrates by statutes enacted in the last 15 years. This particular extension is not particularly significant – it may be appropriate however to require that the applications be heard by a specific pool of Magistrates.[60] The reason for entrusting the matter to the magistrate's courts lie in desiring to make the process accessible to people and cost-effective, furthermore magistrates are entrusted with a certain responsibility regarding the upholding of the peace.[61]

The underlying principles of the proposal mean that it has application beyond what is ordinarily considered to be a "political demonstration".

12 Innovative Demonstrations

A group of students forming an environmental society at Rhodes University embarked in 2008 on a gathering concerning the failings of Makana environmental policy (and policy implementation) in so far as the requirements of the ideal of Environmental Justice are concerned. The gathering was a street clean up march, in which gatherers collected refuse along a path in the under-catered eastern portion of central Grahamstown before placing the sealed refuse bags in front of City Hall, Grahamstown – in a manner that did not block any entrances. This particular gathering was perfectly licit due to the fact that the Municipal by-laws pertaining to dumping created the necessary lee-way by stipulating that any person who dumped any material in the area of the municipality would be required to remove such material within 7 days. Should the Municipality rectify this loop-hole or demonstrators in other municipalities find themselves with municipal by-laws that are less

59 In contrast to the South African situation in which there appears, on a preliminary search, to be no legal literature whatsoever – the Todd article cited is a very short informational piece.

60 The simplest approach would be to require the application be heard by regional civil magistrates and to impose the fairly burdensome application procedures set out in the *Civil Practice Directives for the Regional Courts in South Africa* (2010).

61 As justices of the peace; the nature of this charge within our constitutional dispensation is yet to be tested and certainly warrants attention.

forgiving this demonstration would not be possible. The efficacy of the demonstration was undeniable – the demonstrators had plans to remove the sealed bags from the outside of City Hall four days after the gathering, however the next morning (a Saturday) the refuse was removed by the Municipality – presumably they found the presence of refuse bags unsightly.[62]

Demonstrations demanding the removal of dictators and tyrants (such as Robert Mugabe) frequently possess an element of "stirring language" and may even be interpreted, quite reasonably, as inciting hatred for the dictator. However such demonstrations are far removed from any possibility of effecting violence – on the contrary demonstrators are more likely to be accosted violently, for their expression of opposition to such a tyrant than the tyrant is to be – regardless of how peaceful the demonstration is.

Artistic representations and historic re-enactments of episodes of violence similarly do not enjoy any protection under the RGA. It is my belief that provided the correct setting is established by applicants the expression and re-enactment of mankind's darker side is an important part of reconciling our shared humanity.

Finally it would presently not be possible to arrange a nudist gathering to march down High street (of Grahamstown or any other settlement with a High street) and whilst it may be perfectly sensible to ordinarily restrict such a gathering (on grounds of public morality and aesthetics) the present blanket prohibition undeniably restricts the freedom of expression of persons desiring to express their solidarity with the nudist cause. Furthermore a decadal nudist procession along High street may afford new life into the settlement where it is held – it may also serve the valuable social function of discouraging unorganised and illicit nude processions, by various persons in different degrees of intoxication, that occur on the odd occasion. What is significant however, is that the organisers of such a procession should be required to take satisfactory measures before making application to the court for such permission – a notice in the newspapers forewarning the community at the very least.

13 Conclusion

I have argued that whilst the RGA – but not necessarily its implementation – performs an admirable function in regulating gatherings in so far as the prevention of disruptions to the peace, the RGA fails to make any real headway in the progressive realisation of a truly democratic constitutional state. I have proposed that sections 7 and 8 of

62 As an aside the only disturbance during the gathering was the fact that a municipal official attempted to order the demonstrators not to place the bags at the location they had elected. A complaint was laid with the SAPS and after investigation referred to the Control Prosecutor who has declined to prosecute.

the RGA be amended, as whilst the dignity of the courts is of paramount importance, the present provisions of the RGA are not in sync with our constitutional jurisprudence. Amending the RGA such as to give courts greater powers to facilitate free expression and participation would, I submit, be highly beneficial to South Africa.

The role of the judiciary in balancing flexibility and security*

Marius van Staden
LLB LLM
Lecturer, Faculty of Law, University of Johannesburg

OPSOMMING

Die Rol van die Regbank in die Balansering van Buigsaamheid en Sekuriteit

In wat volg word die regskeppende bevoegdhede van die howe ondersoek en drie verskillende maniere gemerk waardeur die howe kan bydra tot die regulering van die arbeidsmarkte. Eerstens, kan die howe nuwe regulering skep deur middel van hul vertolkingsmandaat. Tweedens, kan die howe nuwe regulering skep deur middel van hul remediërende magte na aanleiding van 'n bevinding van ongrondwetlikheid van bestaande reëls en regulasies. Derdens, kan howe ook verdere regulering deur die wetgewer veroorsaak in hul omgang met en hantering van wetgewing. Daar word gelet op die belang van wetgewing in die algemeen maar ook op die besondere belang van arbeidswetgewing ten opsigte van die regulering van die samelewing en die arbeidsmarkte. Verder word die inherente regskeppende funksie van die howe beskryf. Daar word gemerk dat sekere begrippe juis vaag en wyd opgestel is ten einde dit aan die howe oor te laat om die bepalings van sodanige wetgewing te vertolk en inhoud daaraan te gee. Die howe is dus belangrike akteurs in die buigsaamheid/sekerheid debat aangesien hulle reëls en regulasies "skep" deur middel van hulle vertolkingsmandaat. Daar word egter verder aangevoer dat (as gevolg van die openheid van taal) hierdie verskeinsel nie slegs beperk kan word tot daardie gevalle wat as vaag en wyd bestempel kan word nie. Die vertolkingsbenaderings van onderskeidelik die Hoogste Hof van Appèl en die Grondwetlike Hof, ten opsigte van arbeids-aangeleenthede, word ondersoek. Die howe kan egter ook reg skep na aanleiding van 'n bevinding van ongrondwetlikheid ingevolge artikel 172(1)(a) van die Grondwet en kan ook verdere ingrypings deur die wetgewer in die arbeidsmark veroorsaak. Hierdie verskeinsel laat vrae ontstaan oor hoe howe kan bydra tot die buigsaamheid/sekerheid debat en die regulering van die arbeidsmarkte.

"The courts are the capitals of law's empire, and judges are its princes."[1]

* This article is partially based on a paper presented at the International Labour Law and Social Protection Conference held in Johannesburg, South Africa and hosted by the Faculty of Law and the Centre for International and Comparative Labour and Social Security Law (CICLASS) of the University of Johannesburg from 27 to 30 August 2012.

1 Dworkin *Law's Empire* (1986) 273.

1 Introduction

German scholars tend to characterise states according to what is taken as the province of their main activity. There is *der Kriegerstaat, der Rechtsstaat, der Handelstaat, der Polizeistaat* and so on.[2] Seeley[3] has noted that we live in a legislation-state, which is a form of state devoted to the business of legislation. Legislation has become an indispensable source of contemporary law, if not the most important source.[4] Since the adoption of the Constitution,[5] parliament has adopted a staggering amount of new acts. Between 1997 and 2012 parliament adopted 825 acts as well as 16 constitutional amendments at the rate of about 56 per year. Parliament also embarked upon an overhaul of the legislative landscape of the labour market and many aspects previously left to the common law, or the parties themselves, are now legislatively regulated.[6] The most important way of addressing deficiencies of the common law effectively is by means of legislation,[7] and the legislature is also an institution that is capable of responding "quickly and effectively to frequently fluctuating circumstances of a socio-economic nature".[8] Labour legislation in South Africa is also described as "allies of the Constitution" and "enjoys a considerable status and has a very special role to play in the fulfilment of crucial constitutional objectives".[9]

Of course labour markets cannot be said to be solely regulated by legislation,[10] but contemporary debates concerning the regulation of labour markets focus primarily on three actors – government, business

2 Seeley *Introduction to Political Science: Two Series of Lectures* (1896) 140. *Der Kriegerstaat* refers to a state organised for war, *der Rechtsstaat* refers a state organised around the principle of the rule of law and individual rights, *der Handelstaat* refers to a state devoted to the advancement of trade and der *Polizeistaat* refers to a police-state.

3 146.

4 Du Plessis "The status and role of legislation in South Africa as a constitutional democracy: some exploratory observations" 2011 *PELJ* 92 93.

5 Constitution of the Republic of South Africa, 1996.

6 They include the Labour Relations Act 66 of 1995; the Basic Conditions of Employment Act 75 of 1997; the Compensation for Occupational Injuries and Diseases Act 130 of 1993; the Unemployment Insurance Act 63 of 2001; the Occupational Health and Safety Act 85 of 1993; the Skills Development Act 97 of 1998; the Public Service Labour Relations Act 105 of 1994; the Education Labour Relations Act 146 of 1993; the Employment Equity Act 55 of 1998 and the various amendments to these pieces of legislation.

7 Basson *et al Essential Labour Law* (2005) 9.

8 *Martin v Murray* 1995 ILJ 589 (C) 601E-H.

9 Du Plessis 2011 *PELJ* 92 95, 97. The LRA was enacted "to give effect to and regulate the fundamental rights conferred by section 27" of the Interim Constitution, but s 27 neither explicitly required nor envisaged legislation amplifying and giving more concrete effect to it. S 23(5) and (6) of the 1996 Constitution do, however, envisage legislation to regulate collective bargaining.

10 The most important ways in which employment relationships are regulated in South Africa are through legislation, collective agreements and the contract of employment – Van Niekerk (ed) *Law@work* (2008) 57.

and labour – and the need espoused by these actors for either more or fewer legislative interventions.[11] The role of the judiciary in contributing to this debate has largely been ignored.[12] This may be attributed to an overly narrow view of the separation of powers doctrine and the role of the courts, wherein the regulation of society is properly held to be exclusively the domain of the legislature.[13] Alternatively, this can be attributed to an overly positivistic view of law where legal change is seen solely as political process.[14] At some stage or another, all law students are taught *iudicis est ius dicere non dare*.[15] In such a view labour law is created by Parliament and applied by the courts. Consider *Du Plessis v De Klerk*, where the Constitutional Court remarked on the function of the courts within a democracy.[16] The court warned that the Constitution does not contemplate a dikastocracy and that the "role of the courts is not effectively to usurp the functions of the legislature". Courts, it was held, "should not establish new, positive rights and remedies on its own". Instead judicial function is described as ensuring that legislation does not violate fundamental rights, interpreting legislation in a manner that furthers the values expressed in the Constitution, and ensuring that common law and custom are developed to harmonise with the Constitution.[17]

This does not mean to say that the judiciary has no legislative function and that it cannot contribute to the regulation of labour markets.[18] Indeed, it has become trite that "[l]aw making is an inherent and essential part of the judicial process".[19] In *Matiso v The Commanding Officer, Port Elizabeth Prison* it was stated that judges "invariably 'create' law" notwithstanding that "[f]or those steeped in the tradition of parliamentary sovereignty, the notion of Judges creating law, and not

11 This is described as the tripartite character of labour law.
12 Davies "Judicial self-restraint in labour law" 2009 *ILJ* 278.
13 Botha *Statutory Interpretation: An Introduction for Students* (2005) 49.
14 Morrison *Jurisprudence: From the Greeks to Post-Modernism* (1997) 423.
15 It is the province of a judge to expound the law, not to make it.
16 1996 3 SA 850 (CC).
17 Par 181.
18 Regulation, in this context, is to be understood as a wider term than legislation. The OECD defines "regulation" as "imposition of rules by government, backed by the use of penalties that are intended specifically to modify the economic behaviour of individuals and firms in the private sector". See http://stats.oecd.org/glossary/detail.asp?ID=3295 (accessed 2012-07-02). As such regulation is understood to be a wider term than legislation although legislation is understood to be regulation. Davies *Perspectives on Labour Law* (2004) 54 describes "modes of regulation in labour law" as the "different ways in which labour law is created and applied".
19 *Zinmat Insurance Co Ltd v Chawanda* 1991 2 SA 825 (ZSC) 832H-I. See Du Plessis *The Interpretation of Statutes* (1986) 143; Labuschagne "Die uitlegvermoede teen staatsgebondenheid"1978 *TRW* 42 62; Labuschagne "Regsdinamika: opmerkings oor die aard van die wetgewingsproses" 1983 *THRHR* 422; Le Roux "Undoing the past through statutory interpretation: the Constitutional Court and marriage laws of apartheid" 2005 *Obiter* 526 537.

merely interpreting and applying the law, is an uncomfortable one".[20] In this article it will be shown that courts can regulate labour markets in three distinct ways. Firstly, the courts can, in effect, create new regulation through their interpretative mandate. Secondly, the courts can create new regulation through their remedial powers, following a finding of unconstitutionality of existing regulation. Finally, courts may also trigger or cause further regulation by the legislature. The purpose of this article is solely to classify and to record this phenomenon and not to pass judgement on it or to comment on the desirability thereof. The article will also not consider other ways in which judges can contribute to the regulation of the labour markets, such as, for example, through the interpretation of collective agreements or the development of the common law, but will instead focus on the way in which judges regulate through interactions with legislation. The realisation that the courts are important actors in the regulation of the labour markets can significantly impact upon our understanding of how labour markets are regulated and as such allow labour lawyers, armed with this realisation, to examine how judges contribute to balancing flexibility and security.

2 Interpretation as Regulation

Labour legislation is often drafted in relatively general terms.[21] It is then left to the courts to interpret the provisions of such legislation and the courts therefore become important actors in the flexibility/security debate as they "create law" or regulate labour through their interpretative mandate.[22] Consider, for example, the vague concept "unfair labour practice" as contained in section 23 of the Constitution.[23] When the concept was first introduced in South Africa it was defined as "any labour practice that in the opinion of the Industrial Court is an unfair labour practice".[24] In effect the old industrial court was therefore given

20 1994 3 SA 592 (SE) 597I-598B. In *Doctors for Life International v Speaker of the National Assembly* 2006 6 SA 416 (CC) the Constitutional Court warned that "[c]ourts must be conscious of the vital limits on judicial authority and the Constitution's design to leave certain matters to other branches of government. They too must observe the constitutional limits of their authority. This means that the judiciary should not interfere in the processes of other branches of government unless to do so is mandated by the Constitution" (par 37). The Court did however stress that separation of powers concerns, although important, "cannot be used to avoid the obligations of a court to prevent the violation of the Constitution" (par 200).

21 *National Education Health and Allied Workers Union v University of Cape Town* 2003 2 BCLR 154 (CC) par 34.

22 Du Plessis *Re-Interpretation of Statutes* (2002) 97-98; Du Plessis (1986) 143; Labuschagne "Die uitlegvermoede teen staatsgebondenheid" 1978 *TRW* 42 62; Labuschagne "Regsdinamika: opmerkings oor die aard van die wetgewingsproses" 1983 *THRHR* 422; Le Roux "Undoing the past through statutory interpretation: the Constitutional Court and marriage laws of apartheid" 2005 *Obiter* 526 537.

23 S 23(1) Constitution reads as follows: "Everyone has the right to fair labour practices."

24 S 1(f) Industrial Conciliation Amendment Act 94 of 1979.

extensive discretion to decide for itself what conduct amounted to unfair labour practices and what did not and this leeway, according to some, "amounted to a licence to legislate".[25] Later interventions by the legislature to introduce more specific definitions could also not produce the intended certainty and produced general and open-ended definitions requiring the court to use its own discretion in interpreting it.[26] In *National Education Health and Allied Workers Union v University of Cape Town,* the Constitutional Court found it "neither necessary nor desirable" to define this concept as it is currently contained in the Constitution.[27] Instead the court preferred the concept to be left to gather meaning within the courts.[28] The definition of "employee" as currently contained in labour legislation may also be described as open-ended.[29] Recall that the protection extended by labour legislation, generally applies only to those persons who are defined as "employees".[30]

In the United Kingdom, labour law is also fraught with terms that may be said to be general or vague.[31] "Reasonableness" in the context of unfair dismissals and implied terms in the contract of employment,[32] and "proportionality" with regards to discrimination law and cases arising under the United Kingdom Human Rights Act, 1998 are two such examples.[33] The definition of "employee" as contained in section 230(1) of the Employment Rights Act, 1996 may also be said to be vague.[34] In Germany the Civil Code (*Bürgerliches Gesetzbuch*), the Protection against

25 Thompson & Benjamin *South African Labour Law* (1997) A-60.

26 Vettori *Alternative Means to Regulate the Employment Relationship in the Changing World of Work* (2005 thesis University of Pretoria) 301 available at http://upetd.up.ac.za/thesis/available/etd-11082005-142503/unrestricted/00front.pdf (accessed 2012-07-02) with reference to Thompson & Benjamin *op cit*. The definition was amended by s 1(h) Labour Relations Act Amendment Act 95 of 1980, s 1(h) Labour Relations Act Amendment Act 83 of 1988 and s 1 Labour Relations Act Amendment Act 9 of 1991.

27 2003 2 BCLR 154 (CC) par 33.

28 Par 34.

29 S 213 LRA defines an "employee" as "(a) any person, excluding an independent contractor, who works for another person or for the State and who receives, or is entitled to receive, any remuneration; and (b) any other person who in any manner assists in carrying on or conducting the business of an employer".

30 Van Niekerk (ed) (2008) 57; Grogan *Workplace Law* (2009) 34. Contrast with s 23 of the Constitution which affords "everyone" protection from unfair labour practices, and every "worker" the right to form and join a trade union, to participate in the activities and programmes of a trade union and to strike. According to Cheadle "Labour relations" in *South African Constitutional Law: The Bill of Rights* (eds Cheadle, Davis and Haysom) (2006) 18-3 the use of the term "everyone" in s 23(1) Constitution should not be interpreted to extend beyond the employment relationship, as the boundaries of the right are circumscribed by the reference to labour practices. In *SA National Defence Union v Minister of Defence* 1999 ILJ 2265 (CC) par 25 the court held that the term "worker" as utilised in s 23(2) Constitution is broader than the concept "employee".

31 Davies 2009 *ILJ* 278.

32 *Iceland Frozen Foods v Jones* 1983 ICR 17.

33 Davies 2009 *ILJ* 278.

34 Refer to *Express and Echo Publications Ltd v Tanton* 1999 IRLR 367 (CA).

Dismissal Act (*Kündigungsschutzgesetz*), and the Works Constitution Act (*Betriebsverfassungsgesetz*) also contain general principles rather than specific rules, inevitably drawing the courts into a creative role as they "create" regulation through their interpretive function.[35]

The creative role of the courts cannot be limited to only those cases in which legislative provisions may be said to be vague or drafted in general terms. The openness of language has always produced a proliferation of meanings and this has only been exacerbated by the so-called linguistic or hermeneutical turn in legal interpretation that has been enhanced by the advent of constitutional democracy in South Africa.[36] The linguistic or hermeneutical turn describes a situation in which "[m]eaning is not discovered in (and retrieved from) a construable text, but is made in dealing with the text".[37] Prior to the adoption of a justiciable constitution in South Africa, statutory interpretation proceeded in terms of the now famous *dictum* in *Venter v R*.[38] In terms of this "golden rule" the aim of interpretation was "to ascertain the intention which the legislature meant to express from the language which it employed".[39] In this very case Innes CJ stated that "no matter how carefully words are chosen there is a difficulty in selecting language which, while on the face of it expressing generally the idea of the framer of the measure, will not, when applied under certain circumstances go beyond it, and when applied under other circumstances fall short of it."[40] Accordingly, as early as 1950 Van Heerden AJ observed in *Sachs v Dönges NO* that judges "while purporting merely to expound and apply the law sometimes make law in the process".[41] Indeed, according to Walker "[h]ardly any form of words can be thought of which is not in some circumstances, ambiguous and requiring interpretation".[42] Donaldson J explained this as follows in the British case *Corocraft v Pan-American Airways*:

> In the performance of this [interpretative] duty the judges do not act as computers into which are fed the statute and the rules for the construction of statutes and from whom issue mathematically correct answers. The interpretation of statutes is a craft as much as a science, and the judges as craftsmen, select and apply the appropriate rules as tools of the trade. They are not legislators but finishers and polishers of legislation which comes to them in a state requiring various degrees of further processing.[43]

35 Berger & Neugart "How German labor courts decide: an econometric case study" 2011 *German Econ R* 56.
36 Du Plessis "Interpretation" in Woolman *et al* (eds) *Constitutional Law of South Africa* (eds Woolman *et al*) (2008) 32-44, 32-84.
37 *Supra*.
38 1907 TS 910.
39 914–915.
40 *Supra*.
41 1950 2 SA 265 (A) 312.
42 *The Oxford Companion to the Law* (1980) 644.
43 1968 2 All ER 1059 1071.

Even the most enthusiastic advocates of positivism therefore concede that the courts have a creative role when legislative intention is unclear from the text of the legislation.[44]

Following the toppling of the notion of the "intention of the legislature" by constitutional supremacy in South Africa "broad" purposive interpretation is slowly supplanting (or has already supplanted, some may claim) the old "golden rule" of statutory interpretation as described above.[45] Section 39(2) of the Constitution states that anyone "[w]hen interpreting any legislation must promote the spirit, purport and objects of the Bill of Rights". Botha declares that "[t]he fundamental principle of statutory interpretation is that the purpose of the legislation must be determined in the light of the spirit, purport and objects of the Bill of Rights in the Constitution". It is striking that this principle endorses the purposivist approach whilst qualifying it at the same time.[46] This method of statutory interpretation is generally referred to as "teleological interpretation",[47] a "value-activating strategy",[48] or the "value-coherent theory" of statutory interpretation.[49] It has become commonplace in South Africa for this principle to guide the interpretation of legislation and has been endorsed by the Constitutional Court,[50] although reliance is still regularly placed on other theories of statutory interpretation.[51] It

44 Dias *Jurisprudence* (1976) 464-466.
45 Du Plessis (2008) 32–52.
46 Botha (2005) 10, 66, 75.
47 Botha (2005) 59. The word "teleological" is derived from the Greek word "telos" meaning "[t]he end of a goal-oriented process" – see www.the freedictionary.com/telos (accessed 2012-07-10).
48 Devenish *Interpretation of Statutes* (1992) 40.
49 Devenish (1992) 39.
50 *African Christian Democratic Party v Electoral Commission* 2006 3 SA 305 (CC); *Department of Land Affairs v Goedgelegen Tropical Fruits (Pty) Ltd* 2007 6 SA 199 (CC).
51 The Supreme Court of Appeal has continued to rely on the old "golden rule" of statutory interpretation. Refer to *Randburg Town Council v Kerksay Investments (Pty) Ltd* 1998 1 SA 98 (SCA) 107A-B; *Public Carriers Association v Toll Road Concessionaries (Pty) Ltd* 1990 1 SA 925 (A) 942I-J; *Manyasha v Minister of Law and Order* 1999 2 SA 179 (SCA) 185B-C; *Commissioner, SA Revenue Service v Executor, Frith's Estate* 2001 2 SA 261 (SCA) 273G-I; *Bastian Financial Services v General Hendrik Schoeman Primary School* 2008 4 All SA 117 (SCA) par 16 (although the court also endorsed interpreting provisions contextually). In *Bato Star Fishing (Pty) Ltd v Minister of Environmental Affairs and Tourism* 2004 4 SA 490 (CC) the Constitutional Court rejected the Supreme Court of Appeal's continued reliance on literalism as is the command of s 39(2) of the Constitution that every court "must promote the spirit, purport and objects of the Bill of Rights" when interpreting any legislation and that "[i]mplicit in this command are two propositions: first, the interpretation that is placed upon a statute must, where possible, be one that would advance at least an identifiable value enshrined in the Bill of Rights; and second, the statute must be reasonably capable of such interpretation" (par 72). In addition it was held that "[t]he emerging trend in statutory construction is to have regard to the context in which the words occur, even where the words to be construed are clear and unambiguous" (par 90).

should be noted that some have lamented this development.[52] The approach was best described (although not explicitly endorsed) in *African Christian Democratic Party v Electoral Commission*:

> [I]n approaching the interpretation of provisions of ... legislation, courts ... must understand those provisions in the light of their legislative purpose within the overall ... [legislative] framework. That framework must be understood in the light of the important constitutional rights and values that are relevant.[53]

According to Le Roux this:

> [b]roader [read purposive] approach which the Court favours includes the following distinct steps: (i) establish the central purpose of the provision in question; (ii) establish whether that purpose would be obstructed by a literal interpretation of the provision; if so, (iii) adopt an alternative interpretation of the provision that 'understands' [read promotes] its central purpose; and (iv) ensure that the purposive reading of the legislative provision also promotes the object, purport and spirit of the Bill of Rights.[54]

The Constitutional Court has also expressly endorsed this approach with regards to labour matters in *National Education Health and Allied Workers Union v University of Cape Town* (*NEHAWU*)[55] and most recently in *Aviation Union of South Africa v South African Airways (Pty) Ltd* (*Aviation Union*).[56] These cases ironically considered the interpretation of the very same section of the Labour Relations Act (LRA): section 197.[57] The

52 Le Roux "Directory provisions, section 39(2) of the Constitution and the ontology of statutory law: *African Christian Democratic Party v Electoral Commission* 2006 (3) SA 305 (CC)" 2006 *SAPL* 382 384: The author argues that the court in *African Christian Democratic Party* "too quickly sacrificed the formal integrity or letter of the legislative text in favour of an overconfident appeal to its organic or dynamic substance, purpose or spirit". Refer to Fagan A "In defence of the obvious – ordinary language and the identification of constitutional rules" 1995 *SAJHR* 545; and Fagan E "The longest erratum note in history" 1996 *SAJHR* 79; and the response thereto by Davis "The twist of language and the two Fagans: please sir may I have some more literalism!" 1996 *SAJHR* 504. In addition some have also argued that purposive interpretation has to potential of turning into a rather unruly horse. Refer in general to Du Plessis (2008) 32-54, 32-56.

53 2006 3 SA 305 (CC) par 34. The court therefore concluded that "[t]he purpose of section 14 ... is to ensure that a deposit is paid by a political party [so as] to establish that they have a serious intention of contesting the election. There is no central legislative purpose attached to the precise place where the deposit is to be paid. In my view, to interpret section 14 ... in a manner which prohibits [payment in Pretoria] would be to read the provision unduly narrowly and to misunderstand its central purpose" (par 27).

54 Le Roux 2006 *SAPL* 382 386.

55 Par 62.

56 2012 1 SA 321 (CC).

57 The facts of the case may be summarised as follows: SAA transferred that part of its business concerned with facilities management to LGM. Seven years later, SAA cancelled the agreement and notified LGM that it had to prepare a hand-over plan. SAA had the right to buy back or re-acquire all assets and inventory and the use of all property that had been conferred on

purpose of section 197 of the LRA is to protect the employment of workers and to facilitate the sale of businesses as going concerns by enabling the new employer to take over the workers.[58] This section has caused challenging interpretative difficulties for South African courts.[59] Section 197(1)(b) of the LRA defines "transfer" as follows: "the transfer of a business *by* one employer ('the old employer') to another employer ('the new employer') as a going concern."[60] In *NEHAWU* the Constitutional Court held that the interpretation of section 197 of the LRA in particular and labour matters in general was to proceed as follows:

> The proper approach to the construction of section 197 is to construe the section as a whole and in the light of its purpose and the context in which it appears in the LRA. In addition, regard must be had to the declared purpose of the LRA to promote economic development, social justice and labour peace. The purpose of protecting workers against loss of employment must be met in substance as well as in form. And, as pointed out earlier, it also serves to facilitate the transfer of businesses.[61]

It is generally accepted that section 197 applies to most cases of outsourcing, but the application of the section to second-generation outsourcing has been more problematic precisely because of the definition of transfer as contained in section 197(1)(b) of the LRA and the use of "by" therein.[62] This problem was perhaps best explained in *COSAWU v Zikhethele Trade (Pty) Ltd*:

> A mechanical application of the literal meaning of the word 'by' in section 197(1)(b) would lead to the anomaly that workers transferred as part of first generation contracting-out would be protected whereas those in second generation scheme would not be, when both are equally needful and deserving of the protection. The possibility for abuse and circumvention of the statutory protections by unscrupulous employers is easy to imagine. … I am in agreement ... that section 197(1)(b) might be better interpreted to apply to transfers 'from' one employer to another, as opposed to only those effected 'by' the old employer.[63]

---

57 LGM when the lease was entered into. SAA contended that this would not be a transfer of business within the meaning of the LRA and that it was not obliged to take over the LGM employees involved in the business that had originally been transferred.

58 *National Education Health and Allied Workers Union v University of Cape Town* 2003 2 BCLR 154 (CC) par 54.

59 Refer to *South African Municipal Workers' Union v Rand Airport Management Co Ltd* 2005 3 BLLR 241 (LAC); *Crossroads Distributions (Pty) Ltd t/a Jowells Transport v Clover SA (Pty) Ltd* 2008 6 BLLR 565 (LC); *Chemical Energy Paper Printing Wood & Allied Workers Union v Print Tech (Pty)* Ltd 2010 31 ILJ 1850 (LC); *Zikhethele Trade (Pty) Ltd v COSAWU* 2008 2 BLLR 163 (LAC).

60 Emphasis added.

61 Par 62.

62 Van Niekerk (2008) 335: "Second-generation contracting occurs when a new contractor (which may but not necessarily have been the service provider to whom a business function was initially outsourced) replaces the incumbent contractor."

63 2005 26 ILJ 1056 (LC) par 29.

In *Aviation Union* the Supreme Court of Appeal was unwilling to accept that such "abuse" (as the court termed it) of the plain meaning of the section was warranted.[64] The approach of the Supreme Court of Appeal was decidedly reminiscent of outdated modes of statutory interpretation which has been debunked in South Africa. The following *dictum* serves as a textbook example of the old "golden rule" of statutory interpretation with its roots in both literalism and intentionalism:

> The choice of language in section 197 is plain and unambiguous. By the deliberate use of the word 'by', the legislature showed that it intended section 197 to apply to a situation where there are at least two positive actors in the process. The ordinary meaning of the word 'by' requires positive action from the old employer who transfers the business to the new employer.[65]

What is surprising here is that the Supreme Court of Appeal shows not only contempt for contemporary developments in statutory interpretation in South Africa, but also for established law on how this very section of the LRA was to be interpreted as enunciated in *NEHAWU*.[66] This is not to say that the role of text becomes suddenly irrelevant and the Supreme Court of Appeal might have been right in its refusal to substitute the meaning of one word with that of another. As Rosenau puts it, "[t]he reader may construct the text, but the text in turn controls the encounter".[67] The difficulty here relates to the amount of control that the text exercises upon a construction. It is perhaps for this reason that both the majority of the Constitutional Court per Yacoob J and the minority per Jafta J held in *Aviation Union* that it is "unnecessary to equate the word 'by' with 'from'".[68] But even though the court could not read "by" to mean "from" it would be wrong to categorise its mode of interpretation as literalist. In fact the court held that section 197 of the LRA does apply to second generation outsourcing. What the court does is to read the provision purposively as they are obligated to do in terms of *NEHAWU* but also within the context of the whole provision:

> Determining the operation of the section with reference to a single word is not the correct approach to its interpretation. The whole section must be read in its proper context. Reading section 197 as a whole in the context of where

64 *South African Airways (Pty) Ltd v Aviation Union of South Africa* 2011 3 SA 148 (SCA) par 32.

65 Par 31.

66 Two further points of criticism can also be noted against the approach of the SCA. Firstly, it may be said that the SCA commits the error of "dis-integration" where the court "turns a blind eye to the systematic interconnectedness of text-components and tries to understand them in splendid isolation from one another". Additionally the SCA may be criticised for its excessive peering at an individual word or words - See Du Plessis "Interpretation of Statutes and the Constitution" in *Bill of Rights Compendium* (2012) par 2C37, 2C40. It may however be argued that these points are logical manifestations of the courts continued reliance on literalism and intentionalism.

67 *Post-Modernism and the Social Sciences: Insights, Inroads, and Intrusions* (1992) 25.

68 Par 81.

> it is located in the LRA and paying sufficient attention to its purpose and the objects of the LRA, reveal that it applies to any transaction that transfers a business as a going concern.[69]

In the process the court attached a meaning to the provision which may not have been the first meaning that springs to mind when reading the provision, but the words are still reasonably capable of bearing that meaning.[70]

The Supreme Court of Appeal has recently had occasion to rethink its continued reliance upon outdated interpretive modes in *Natal Joint Municipal Pension Fund v Endumeni Municipality*.[71] "Interpretation", said the court:

> [i]s the process of attributing meaning to the words used … having regard to the context provided by reading the particular provision or provisions in the light of the document as a whole and the circumstances attendant upon its coming into existence.[72]

The court also showed why continued deference to the intention of the legislature was inappropriate.[73] The court acknowledged that "[m]ost words can bear several different meanings or shades of meaning" and that the idea of an ordinary grammatical meaning of words is therefore a misnomer.[74]

3 Remedies as Regulation

Prior to the advent of constitutional democracy in South Africa, no judicial review of legislation was possible except to the extent that delegated legislation was capable of review against the standards of administrative law.[75] Section 172(1)(a) sets out the powers of courts in constitutional matters and maintains that "[w]hen deciding a constitutional matter within its power, a court *must* declare that any law or conduct that is inconsistent with the Constitution is invalid to the extent of its inconsistency".[76]

---

69 Par 55.

70 By this means the court can be said to pass the test of O'Regan J in her dissenting judgement in *South African Police Services v Public Servants Association* 2007 5 BLLR 383 (CC) par 94.

71 2012 2 All SA 262 (SCA).

72 Parr 18, 19: "This is the approach that courts in South Africa should now follow, without the need to cite authorities from an earlier era that are not necessarily consistent and frequently reflect an approach to interpretation that is no longer appropriate."

73 Parr 20-24.

74 Par 25.

75 *Fedsure Life Assurance v Greater Johannesburg Transitional Metropolitan Council* 1999 1 SA 374 (CC) parr 27, 28.

76 Emphasis added. Compare to s 4 United Kingdom Human Rights Act, 1998, which provides that a court "may make" a declaration of incompatibility when the Court is satisfied that a legislative provision cannot be read and

In *Van Rooyen v The State*[77] the Constitutional Court explained how constitutional challenges to legislation must be dealt with. Firstly, all law must be interpreted ("read down") to avoid inconsistency with the Constitution. Secondly, if this is not possible the law (or at least the offending parts thereof) must be declared unconstitutional and invalid. Thirdly, the court may remedy the unconstitutionality by means of severance, reading in or notional severance. If this is not possible then the ultimate remedy must be granted in declaring the law (completely) invalid. In addition, the Constitution also empowers the courts to limit the impact of the order of invalidity by suspending or limiting the retrospective effect of the order.[78] The Constitutional Court has also, on several occasions, allowed an interim remedy of notional or actual severance, or reading in during a period of suspension to diminish the effects of the continued violation of rights.[79]

A declaration of invalidity to the extent of the inconsistency is the default remedy following a finding of inconsistency and will only be departed from if a more limited order (such as reading in, notional or actual severance, suspending or limiting the retrospective effect of the order, or a combination of these) will provide a better outcome.[80] There

given effect to in a manner compatible with the European Convention on Human Rights.

77 2002 5 SA 246 (CC) par 88.

78 S 172(1)(b).

79 *Dawood v Minister of Home Affairs* 2000 3 SA 936 (CC); *Janse van Rensburg v Minister of Trade and Industry* 2001 1 SA 29 (CC); *Booysen v Minister of Home Affairs* 2001 4 SA 485 (CC); *Volks v Robinsion* 2005 5 BCLR 446 (CC); *Moseneke v The Master of the High Court* 2001 2 SA 18 (CC); *South African Liquor Traders Association v Chairperson, Gauteng Liquor Board* 2006 8 BCLR 901 (CC); *C v Department of Health and Social Development* case no 55/11 (CC) (unreported). Note however that in *J v Director General, Department of Home Affairs* 2003 5 SA 621 (CC) par 22 it was held that "[w]here the appropriate remedy is reading in words [or severance or notional severance] in order to cure the constitutional invalidity of a statutory provision, it is difficult to think of an occasion when it would be appropriate to suspend such an order. This is so because the effect of reading in is to cure a constitutional deficiency in the impugned legislation. If reading in words does not cure the unconstitutionality, it will ordinarily not be an appropriate remedy. Where the unconstitutionality is cured, there would usually be no reason to deprive the applicants or any other persons of the benefit of such an order by suspending it. Moreover the legislature need not be given an opportunity to remedy the defect, which has by definition been cured. In the present case, the effect of the order is not to leave a lacuna but to remedy the constitutional defect complained of by the applicants by a combination of reading in and striking down. Under the circumstances, it is not an appropriate case for our order to be suspended." Bishop "Remedies" in *Constitutional law of South Africa* (eds Woolman *et al*) (2008) 9-125, defends the practice on the ground that, when a court awards this hybrid remedy it has already concluded that reading in, notional or actual severance is an inappropriate permanent measure but requires a "stop-gap measure" until the legislature gets around to deciding how to cure the defect.

80 Currie & De Waal *The Bill of Rights Handbook* (2005) 199.

are cases where immediate and fully retrospective invalidity was found to be the only workable remedy.[81]

Reading in is predominantly used when the inconsistency is caused by an omission, and it is necessary to add words into the statutory provision to cure it.[82] Severance (or actual severance) is the corollary remedy in terms of section 172 of the Constitution. Section 172(1)(a) provides that a court must declare that any law or conduct that is inconsistent with the Constitution invalid to the extent of its inconsistency. Thus, a court is required to strike down only unconstitutional parts of the law. This may be a particular section, subsection or parts thereof.[83] Notional severance, like actual severance, involves removing offending parts from the legislative text whilst leaving other words and provisions intact. No actual words are deleted or removed from the law. "Instead the

81 See for example *S v Makwanyane*1995 3 SA 391 (CC); *Larbi-Odam v Member of the Executive Council for Education (North-West Province)* 1998 1 SA 754 (CC); *Magajane v Chairperson, North West Gambling Board* 2006 5 SA 250 (CC); *Case v Minister of Safety and Security, Curtis v Minister of Safety and Security* 1996 3 SA 617 (CC). Note that, although in terms of s172(2)(a) any court may declare a statute invalid, declarations of invalidity of national or provincial legislation have no force unless they are confirmed by the Constitutional Court. Constitutional court rule 16(4) provides: "A person or organ of state entitled to do so and desirous of applying for the confirmation of an order in terms of section 172(2)(d) of the Constitution shall, within 15 days of the making of such order, lodge an application for such confirmation with the Registrar and a copy thereof with the Registrar of the court which made the order, whereupon the matter shall be disposed of in accordance with directions given by the Chief Justice."

82 Currie & De Waal (2005) 204. In *National Coalition for Gay and Lesbian Equality v Minister of Home Affairs* 2000 2 SA 1 (CC) par 75 the Constitutional Court laid down the following requirements which should be considered before this remedy is granted: "In deciding to read words into a statute, a court should also bear in mind that it will not be appropriate to read words in, unless in so doing a court can define with sufficient precision how the statute ought to be extended in order to comply with the Constitution. Moreover, when reading in (as when severing) a court should endeavour to be as faithful as possible to the legislative scheme within the constraints of the Constitution. Even where the remedy of reading in is otherwise justified, it ought not to be granted where it would result in an unsupportable budgetary intrusion. In determining the scope of the budgetary intrusion, it will be necessary to consider the relative size of the group which the reading in would add to the group already enjoying the benefits. Where reading in would, by expanding the group of persons protected, sustain a policy of long standing or one that is constitutionally encouraged, it should be preferred to one removing the protection completely."

83 Currie & De Waal (2005) 200-210. *Coetzee v Government of the Republic of South Africa, Matiso v Commanding Officer Port Elizabeth* 1995 4 SA 631 (CC) par 16 qualified the use of the remedy as follows: "Although severability in the context of constitutional law may often require special treatment, in the present case the trite test can properly be applied: if the good is not dependent on the bad and can be separated from it, one gives effect to the good that remains after the separation if it still gives effect to the main objective of the statute. The test has two parts: first, is it possible to sever the invalid provisions and second, if so, is what remains giving effect to the purpose of the legislative scheme?"

unconstitutional reach of the provision is identified by limiting the cases to which the law can apply or by making its valid operation subject to a condition."[84]

When a court utilises one of these remedies it is obvious that they create law and they do so unashamedly because the Constitution requires of them to do so. The Constitutional Court has recently had occasion to remind actors in the debate regarding the appropriateness of judicial law making that the courts are not "the ultimate word on pronouncing on the law" as Parliament may always amend the remedy through its legislative power. "Instead", said the court, "it initiates a conversation between the legislature and the courts."[85]

4 Triggering New Regulation

A recent economic study by Hefeker and Neugart has empirically found that uncertainty in the way the courts interpret labour legislation will cause government to opt for more regulation.[86] To demonstrate how the courts may contribute to the adoption of new regulation, the link between inflexible labour markets and unemployment will be considered. Of course it is well established that this link exists.[87] The authors' findings may be summarised as follows.[88] if government was the only internal actor it would decide upon an appropriate trade-off between flexibility and security that it deems appropriate to protect workers and create jobs, and regulate the market as such. Political consequence would be its own. However, government is not the only actor who can regulate the market, as the courts also possess a limited law making function, as described above, and this influences government's decision to adopt new regulations or not. Governments will, for example, not intervene in the labour market to provide security for workers and be willing to take on the political cost of this decision if it can be certain that the courts will do the other part of the job in providing security for workers. If, conversely, government is sure that the courts will tend to favour job creation, government will not embark upon costly reforms to create flexibility in the labour market, but will leave this to the courts. By leaving it to the court to take the necessary steps, there will be no need for the government to take the blame for unpopular policy measures. Therefore, governments will be more likely to intervene in the labour market if they are confronted with a judicial system where the courts are vested with the discretion to follow their own preferences in interpreting the provisions in labour legislation.

84 Bishop (2008) 9-102. Refer to *Ferreira v Levin*, *Vryenhoek v Powell* 1996 1 SA 984 (CC); *Islamic Unity Convention v Minister of Telecommunications* 2008 3 SA 383 (CC).
85 *C v Department of Health and Social Development* 2012 2 SA 208 (CC) par 57.
86 "Labor market regulation and the legal system" 2010 *Int Review Law and Econ* 218 255.
87 World Bank *Doing Business in 2006 – Creating Jobs* (2006).
88 Hefeker & Neugart 2010 *Int Review Law and Econ* 218.

In a sense, it may be argued that lawyers have always been aware of this phenomenon. The abolition of the death penalty in South Africa by the Constitutional Court in *S v Makwanyane*, for example came at a time when it was accepted that the majority of South African's were against the abolition thereof.[89] In allowing this decision to be taken by the Constitutional Court, government could avoid what would have been a costly political decision. In the same way it is also clear that inconsistency in the way that labour courts and tribunals have interpreted certain provisions have caused the legislature to act.[90] On the other hand consistency in the way other provisions were interpreted has not necessitated any costly political intervention by parliament.[91] The question that arises is if the approach of the South African judiciary can be described as "uncertain" and one in terms of which "the courts are vested with the discretion to follow their own preferences in interpreting the provisions in labour legislation". Perhaps this is true in the sense that the Supreme Court of Appeal has continued to rely on certain modes of interpretation whilst the Constitutional Court has relied on others such in the case of *Aviation Union.*[92] But neither of these approaches may be said to be "certain" precisely because the nature of language is not certain. It can also not be said that a court follows its "own preferences" when it interprets a provision purposively in the light of the values enshrined in the Constitution. In any event it may be argued that that which is certain to labour lawyers might seem less so to labour economists and the point should rather be that the judiciary contributes to the regulation of the labour market not just when it interprets legislation but also in how it interprets those provisions.

5 Conclusion

A claim of judicial law-making in judicial dealing with legislation is by its very nature controversial. It moves against some of the basic tenants the legal profession holds dear. Sir Otto Kahn-Freund, in reflecting upon the desirability of the inclusion of labour rights within a justiciable

89 1995 3 SA 391 (CC) par 88.

90 It may, for example, be argued that the inconsistency with which the South African labour courts and tribunals have dealt with temporary employment services has forced parliament's hand as it necessitated legislative intervention. The judiciary was not effective in providing sufficient protection to workers employed by temporary employment services notwithstanding the fact that these workers were also subject to the constitutional guarantees envisaged in s 23 Constitution. Refer to *Vilane/ SITA (Pty) Ltd* 2008 5 BALR 486 (CCMA).

91 It may also be argued that the judiciary's protection of illegal workers such as in the case of illegal immigrants and sex-workers has freed up parliament from taking costly political decisions as to the formal regulation of these workers. Refer to *Kylie v Commission for Conciliation Mediation and Arbitration* 2010 4 SA 383 (LAC); *Discovery Health Limited v Commission For Conciliation, Mediation and Arbitration* 2008 7 BLLR 633 (LC).

92 *South African Airways (Pty) Ltd v Aviation Union of South Africa* 2011 3 SA 148 (SCA).

constitution, warned of the possibility that judges would "read their own notions of policy into the bill of rights … shifting the function of law reform from Parliament … to the Bench and to the Bar".[93] But, as Smit records, "the judiciary has through its power of interpretation the ability to substantially increase the rights, protection and freedom of people".[94] Perhaps the time has come for labour lawyers to realise that, as the Constitutional Court explained in *NEHAWU*, "the courts and the legislature act in partnership to give life to constitutional rights".[95] In any event, a refusal to admit that courts do contribute to the regulation of the labour markets is not helpful. The realisation that courts impact directly upon the regulation thereof allows us to re-examine question of labour law and to re-assess traditional responses to these questions. Responses to the changing nature of work cannot come solely from government, business and labour. If the courts can be seen to be important actors in the regulation of the labour markets then labour lawyers must also ask how the judiciary may be part of the solution to such problems. If a labour market is said to be overregulated, then it must be questioned how the judiciary has contributed to such a situation. As pressure from neo-economists and their political followers to deregulate increases, courts should not be allowed to stand on the side-lines of questions and debates, but, in embracing its creative role, must actively contribute to workable solutions to these pressures.

93 Kahn-Freund "The impact of constitutions on labour law" 1976 *Camb LJ* 270. See also Beatty "Constitutional labour rights: pros and cons"1993 *ILJ* 1.
94 Smit "Towards social justice: an elusive and challenging endeavour" 2010 *TSAR* 1 11.
95 Par 14.

Pashukanis on crime and punishment

Raymond Koen
LLB LLM PhD
Associate Professor, Department of Criminal Justice and Procedure, University of the Western Cape

OPSOMMING

Pashukanis oor Misdaad en Straf

Evgeny Pashukanis is bekend vir die sogenaamde kommoditeits-vorm regsteorie. Die teorie is gebaseer rondom die verhouding tussen die regsvorm en die kommoditeitsvorm, wat wentel om die beginsel van ekwivalensie. Pashukanis het dit as 'n algemene teorie ontwikkel en hy het dit nog altyd as sulks aanvaar.

Sedert die publikasie daarvan in 1924, was die Pashukanis regsteorie die teiken van volgehoue kritiek. Een van die mees prominente punte van kritiek is die bewering dat die teorie nie algemeen toepassbaar is nie en dat die teorie nie van sy historiese oorsprong in kontraktereg kan ontsnap nie. Terwyl die beginsel van ekwivalensie toegepas kan word in privaatreg, weerspieël dit nie die verhoudings in die publiekreg nie. Pashukanis se poging om misdaad en straf as 'n integrale aspek van die kommoditeits-vorm teorie aan te bied, is nie oortuigend nie. Die teorie kan slegs slaag as 'n teorie van privaatreg, maar nie as 'n algemene teorie van die reg nie.

Hierdie artikel beoog om die argument dat Pashukanis se teorie nie 'n algemene teorie is nie, te ondersoek. Dit word gedoen met spesifieke verwysings na die ontleding van misdaad en straf, en poog om te wys, histories en analities, dat die beginsel van ekwivalensie ook in hierdie gebiede van toepassing is. Die ontleding sluit in pleitonderhandeling, borgtog, noodweer, toestemming en toerekeningsvatbaarheid. In die algemeen word die onderlinge verhoudings tussen kontrakte, delikte en misdaad ook ondersoek. Die impak wat die beginsel van ekwivalensie op straf het word ook verduidelik. Ten slotte is die gevolgtrekking dat daar geen teenstrydighede tussen strafreg en die beginsel van ekwivalensie is nie. Inteendeel, strafreg is 'n vertakking van die reg waar hierdie regsvorm die opvallendste is. Die Pashukanis-regsteorie kan dus beskou word as 'n algemene regsteorie.

1 Introduction

Evgeny Pashukanis deservedly is famous as the author of the so-called commodity form theory of law. In his *Law and Marxism* he postulated that the form of legal relations held the key to the Marxist critique of law and that, in turn, the key to comprehending the legal form lay in its relation to the commodity form. The crucial concept here is the principle of equivalence or the equality postulate, which Pashukanis classifies variously as the "first truly juridical idea"[1] or the "juridical soul" of

1 Pashukanis *Law and Marxism: A General Theory* (1978) 168.

criminal proceedings.[2] Just as commodity exchange pivots upon mutual recognition by commodity owners of one another as equals, so legal exchange stipulates reciprocal acceptance by legal subjects of one another as compeers. Indeed, juridification is the *alter ego* of commodification, in that the evolution of the legal form tracks the evolution of the commodity form. In a word, Pashukanis theorised the legal form as the homologue of the commodity form, with both delimited in terms of the principle of equivalence.

Pashukanis made it clear always that the historical genesis of his general theory lay in private law, specifically the law of contract. It is the branch of law which is both the historical and logical repository of the notion of equivalence.[3] By contrast, criminal justice appears to be far removed from the commodity form and the idea of equivalence. This article investigates the relationship between Pashukanism and criminal justice, attempting to prove that the private law derivation of the commodity form theory does not preclude its extrapolation to public law in general and to criminal law in particular. It seeks to convince that the disjunction between Pashukanism and criminal justice is more apparent than real. Pashukanis formulated the commodity form theory as a general theory of law, and the argument herein thus may be read as a defence of that generality.[4]

2 Crime and the Principle of Equivalence

The principle of equivalence, as the centrepiece of the Pashukanist general theory of law grounds the concept of justice in all social formations structured by the commodity economy. It thus also governs the comprehension of criminal justice. However, it is not unusual for commentators to espy a deficiency pertaining to criminal justice in Pashukanis's theory. They submit that it is, at bottom, a theory of private law and, as such, unable to account for criminal law. This position is exemplified by Warrington:

> Pashukanis's theory is really concerned with private law and the chapter on criminal law is only added to attempt a spurious theoretical consistency. Pashukanis merely tries to apply his commodity form theory which had a certain logical force for private law, to criminal law, where in the formulation of Pashukanis at least, it clearly has no place.[5]

2 Pashukanis 177.
3 See Pashukanis 121.
4 See Pashukanis 40. See also Lipson "Is there a Marxist Theory of Law? Comments on Tushnet" in *Marxism: NOMOS XXVI* (eds Pennock & Chapman) (1983) 192: "He maintained … that law as a general form – not merely piece by piece, but as a general form – was linked in history to that economic relationship that he said Marx said was at the bottom of all societies that obtained in the interval between the end of primitive family subsistence and the beginning of true socialism: namely, the relationship of commodity exchange".
5 Warrington "Pashukanis and the Commodity Form Theory" in *Legality, Ideology and the State* (ed Sugarman) (1983) 62.

For Warrington, criminal law "fits uncomfortably into Pashukanis's project".[6] Hirst, too, is unconvinced by the Pashukanist approach to criminal law. He comments:

> Crime in capitalism is conceived on the analogy with private law as the violation by the criminal of a right borne by society, a violation of obligation which requires recompense. But this extension of the legal form is a mere ideological cover.[7]

According to Hirst, the relationship between the commodity form theory and criminal law is bedevilled by "Pashukanis's crudities",[8] which include the "vulgar Marxist-Leninist conception of the state as a coercive apparatus in the hands of the exploiting class".[9] In other words, Pashukanis comprehends the criminal law as an instrument of class domination and his incorporation of it into the legal form is unable to conceal his alleged instrumentalism. Hirst, it seems, would concur with Warrington that Pashukanis's approach to criminal law involves theoretical artifice to generalise his theory beyond its origins in private law.

The scepticism regarding the compass of the commodity form theory of law amounts to a serious assault upon the coherence of Pashukanism. It is necessary, therefore, to examine what Pashukanis has to say on the matter. For him, crime is a special form of contract, a retrospectively imposed contract.

"Felony can be seen as a particular variant of circulation, in which the exchange relation, that is the contractual relation, is determined retrospectively, after arbitrary action by one of the parties."[10]

The party who has taken the arbitrary action is the offender. He wishes a one-sided relation, from which he is the sole beneficiary,[11] with the victim as utter loser.[12] The criminal law intervenes to abolish the privilege of asymmetry claimed by the offender. The state impresses him into a contract after he already has had his satisfaction, and he is forced to render performance to his victim. By his crime, the offender has

6 Warrington 62.
7 Hirst *On Law and Ideology* (1979) 115.
8 Hirst "Introduction" in *Ownership of the Image: Elements for a Marxist Theory of Law* (Edelman) (1979) 10.
9 Hirst *On Law and Ideology* (1979) 114.
10 Pashukanis 168. Melossi & Pavarini *The Prison and the Factory: Origins of the Penitentiary System* (1981) 2 refer to this formulation as "the famous thesis of Pashukanis".
11 See Ripstein *Equality, Responsibility and the Law* (1999) 10: "Crime consists in the pursuit of private rationality in the face of the rights of others, of the wrongdoer's substitution of his private rationality for public terms of reasonableness".
12 See Ripstein 144: "The criminal, by intentionally or recklessly violating the victim's rights, expresses a denial of the victim's value. In substituting private rationality for public standards of reasonableness, the criminal declares the victim's rights irrelevant".

violated the principle of equivalence which defines all things juridical. The criminal law exists to reinstate this juridical prime directive.

The notion of founding the comprehension of crime upon the principle of equivalence has an ancient provenance. As Pashukanis observes, it was Aristotle who was the progenitor of the "definition of crime as an involuntarily concluded contract".[13] This definition constitutes one aspect of the Aristotelian category of remedial or corrective justice which is "manifested in the adjustment of balance in transactions between man and man" (as opposed to distributive justice which is concerned with the "distributions of honour, wealth or whatever else is divisible among those who enjoy citizen rights").[14] Aristotle divided remedial justice into voluntary and involuntary transactions, the former referring to "transactions based on a contract between the parties, which one or the other has broken", and the latter to "transactions independent of any consent, in which one party has wronged the other".[15]

Aristotle's domain of involuntary transactions includes crimes as non-consensual contracts.[16] The offender has trashed the normal relationship of equality with his victim, and claimed a position of superiority for himself. The aim of remedial justice is "to restore a violated and interrupted equality".[17] Aristotle understands justice as "a sort of equality" as opposed to injustice which is a "sort of inequality".[18] He argues thus:

> It makes no difference whether a good man has defrauded a bad man or *vice versa*, nor whether adultery has been committed by a good or a bad man. If one person is in the wrong and another is suffering wrong, ie if one has inflicted and the other sustained an injury, the law looks only to the specific nature of that injury. This kind of injustice, therefore, is an inequality, and the judge tries to equalise it. Even where one person has been wounded, or has suffered death, at another's hands the 'being done to' and the 'doing' are represented by a line divided into unequal segments; but the judge tries by means of the penalty he imposes to equalise the wrong suffered and the wrong done, subtracting from the 'gain'.[19]

In Pashukanist terms, Aristotelian equalisation is achieved by the retrospective construction of the crime as a contract in order to

13 Pashukanis 169.
14 Aristotle (trans J Warrington) *Ethics* (1963) 1130b18-1131a6. See also Ross *Aristotle* (1949) 210.
15 Barker *The Politics of Aristotle* (1952) 363. Aristotle 1130b18*ff* identifies two classes of involuntary transactions, namely, clandestine and violent. The former include theft, adultery, poisoning, procuring, enticement of slaves, assassination and false witness; the latter span assault, imprisonment, murder, robbery with violence, mutilation, abuse and insult.
16 See Ross 211; Barker *The Political Thought of Plato and Aristotle* (1959) 343.
17 Barker (1959) 343.
18 Aristotle 1131b36-37.
19 Aristotle 1132a2-25.

rehabilitate the equality postulate between the offender and victim as involuntary contractants.[20]

2 1 Plea Bargaining

The contemporary exemplar of the Aristotelian contractarian approach to crime is the so-called plea bargain. The standard plea bargain involves the accused pleading guilty in exchange for a sentence lighter than that likely to follow a trial conviction.[21] Despite its criminal justice setting, the plea bargain is a contractual transaction, which complies with all the essential requirements of classical contract law.[22] Indeed, the very notion of a bargain entrains a bundle of contractual relations. Scott and Stuntz explain: "Plea bargains are, as the name suggests, *bargains*: it seems natural to argue that they should be regulated and evaluated accordingly."[23]

They submit further that in the United States, plea bargaining is so routine and ubiquitous that it "is not some adjunct to the criminal justice system; it *is* the criminal justice system".[24] It seems that contract is constitutive of criminal justice as a system, at least in the land of the plea bargain.[25]

2 2 Bail

Unlike the USA, most other criminal jurisdictions continue to rely upon the criminal trial as the barometer of liability. However, even in trial-based systems, contractual relations feature prominently in the crucial issue of bail.[26] Van der Berg is explicit:

---

20 See Ripstein 134: "The criminal law serves primarily to protect and vindicate fair terms of interaction".

21 Thus, the standard plea bargain is really a plea and sentence bargain. Variations include pleading guilty to lesser or fewer charges in exchange for sentence reduction or bargaining only about the plea and leaving the sentence to the discretion of the court. The commodified character of the plea bargain is apparent to Sandefur "In Defence of Plea Bargaining" 2003 *Regulation* 31: "The courtroom may not seem like a place for haggling, but that is exactly what it is, in both civil and criminal contexts. A civil defendant can settle his case for a certain sum; a criminal defendant for a certain amount of time".

22 For a thorough discussion and defence of the contractual core of plea bargaining, see generally Scott & Stuntz "Plea Bargaining as Contract" 1992 *Yale LJ* 1910 who, *inter alia*, advise that the courts have developed "a body of contract-based law to regulate the plea bargaining process". See also Sandefur 28.

23 Scott & Stuntz 1910, original emphasis. See also Sandefur 28.

24 Scott & Stuntz 1912, original emphasis.

25 See generally Fisher *Plea Bargaining's Triumph: A History of Plea Bargaining in America* (2003).

26 Caldwell *Criminology* (1965) 350 defines bail pithily as "security furnished to the court for the appearance of the defendant whenever his presence is needed." See also De Haas "Concepts of the Nature of Bail in English and American Criminal Law" 1946 *U Toronto LJ* 387.

> The granting of bail is in the nature of a contract in terms of which the state commits itself to the accused's continued interim freedom once the court has authorised his release, while the accused commits himself to standing trial. It is apparent that the contracting parties are the state and the accused, although the discretion to grant bail vests in the court [27]

He considers that bail should be governed entirely by freedom of contract and argues that "the court's involvement in the bail contract is anomalous and even undesirable, more so where the state and the accused have agreed upon acceptable terms of the bail contract".[28] He would have the court ejected from the process as far as possible and leave the settlement of the quantum and conditions of bail to the contractants. He wishes freedom of contract to be properly constitutive of this pivotal pre-trial dimension: contractants are to be left to their own devices; only if they are unable to reach consensus would the court have to intercede to resolve the impasse. He even proposes that the Criminal Code be amended to this effect.[29]

2 3 Private Defence

Contractual relations and the equality postulate feature also in two popular substantive criminal law defences, namely, private defence and consent. Both are justifications, in the sense that conduct which complies with their criteria is considered not to be unlawful.

One of the requirements of private defence is that the response of the accused to an unlawful attack, upon himself or a third party, be reasonable. This requirement translates into a measure of proportionality or equivalence between the harms on either side of the criminal confrontation. In other words, the violence deployed by the defender should be commensurate, more or less, with that perpetrated or threatened by the attacker. An excessively violent defence is, *per definitionem*, unreasonable and will transform the defender himself into an unlawful attacker.[30]

In Pashukanist terms, the defender who responds reasonably to an unlawful attack has adhered to the principle of equivalence; and, provided the other requirements of the defence have been satisfied, he will not be liable for the harm suffered by the attacker. The reasonableness of the defence operates to refute the assertion of advantage, of "private rationality", entailed in the original attack and to reconstitute the principle of equivalence. Despite its public law-ness, there is a relatively comfortable fit between private defence and Pashukanis's notion of crime: the contractual *leitmotif* of equivalence is embedded in the composition of the defence.

---

27 Van der Berg *Bail: A Practitioner's Guide* (2012) 9.
28 Van der Berg 10.
29 Van der Berg 11.
30 See Burchell *Principles of Criminal Law* (2005) 239-242; Snyman *Criminal Law* (2008) 109-112.

2 4 Consent

Consent is perhaps the prime example of a contractual justification in the criminal law. It is a defence which proceeds from the autonomy of the legal subject, accepting that each has the freedom to transact with his interests as he sees fit, including assenting to their being violated by another. Subject autonomy is extensive and comprehensive, provided its exercise does not conflict with the *boni mores*.[31] The freedom to consent to harm evidently is analogous to freedom of contract. Both are derived from the philosophy of individualism which pervades the commodity economy and both are grounded in the autonomy of the legal subject as the juridical materialisation of the commodity owner.

It is well established that consent to being harmed must be real, that is, it must not be given under duress or extracted by way of a fraud pertaining to a material aspect of the transaction in question. Also, the consent in question must be lawful, and the person consenting must have the capacity to do so.[32] These stipulations are integral also to the consensus which is required for all valid contracts.[33] In other words, there is a discernible contractual dimension to the criminal law defence of consent. At first blush, the equality postulate may appear inapplicable to situations where one person allows another to infringe his rights or interests. However, consent probably is raised most often in cases of sex crimes and medical treatment. In such cases it is not difficult to relate the postulate to notions of sexual gratification or to therapeutic or non-therapeutic satisfaction.

2 5 Criminal Capacity

In South African law, the accused must possess the intellectual apparatus needed to commit a crime. That is, he has to be able to comprehend the distinction between right and wrong, and to modulate his conduct according to this distinction. An accused who, for whatever reason, lacks either of these abilities at the time of the offence, will escape liability for lack of criminal capacity.[34] This absence founds such defences as youth, insanity, provocation, intoxication and non-pathological criminal incapacity.

Contractual capacity or "the competence to create rights and duties by concluding a contract with another person or persons",[35] is the counterpart of criminal capacity. As with crime, contractants must be

31 It usually stops short only at death: in most jurisdictions the legal subject is not free to consent to being murdered. See Snyman 125; Burchell 326.
32 See Snyman 127; Burchell 340-346.
33 See Van der Merwe *et al Contract: General Principles* (1993) 13-14, 73-74, 85-86, 91-92, 139-141.
34 See Snyman 160; Burchell 358.
35 Kuschke "Criminal Capacity" in *The Law of Contract in South Africa* (eds Hutchison & Pretorius) (2012) 150.

endowed with the intellectual wherewithal to enter into contractual relations, and

> the extent to which a person has capacity to contract depends upon his ability to form and express a legally relevant will, which in turn depends upon the ability to appreciate the nature and effect of his or her act.[36]

Contracts purported to be concluded by persons who do not possess the requisite capacity generally are invalid. Thus, for example, *infantes* do not possess the capacity to contract, and contracts concluded by the insane and heavily intoxicated persons may be voided for lack of contractual capacity or voidable for deficient contractual capacity. Given the historical priority of contract over crime, the roots of criminal capacity lie deep in the soil of contractual capacity.

2 6 Contract, Delict and Crime

Examples, however many or compelling, by themselves cannot comprise a finished argument for the Pashukanist position on crime. Still, the examples considered above, in combination, are highly evocative of the contractual core of criminal law, and give a sense that Pashukanis's "famous thesis" perhaps is not as exotic as critics have insinuated. They suggest that his contractual theory of crime well may be a licit derivation from the commodity form theory of law. The sequel will attempt to theorise this proposition by examining briefly the interrelations of contract, crime and delict.

The law of delict probably stands abreast with the law of contract as being quintessentially private and governed by the principle of equivalence.[37] The "contractual" nature of delict is apparent in the plaintiff's claim for recompense for the harm suffered at the hands of the defendant. Indeed, it is acknowledged generally that a complete overlap between contract and delict is possible, in that a breach of contract may amount also to a delict.[38] However, unlike breach of contract, an action in delict does not take place in the context of a pre-existing voluntary agreement. Hutchison explains:

The essential difference between contractual and delictual obligations is that the former are, as a general rule, voluntarily assumed by the parties themselves, whereas the latter are imposed by law, irrespective of the will of the parties.[39]

---

36 Kuschke 150.

37 See Ripstein 246.

38 See Boberg *The Law of Delict: Volume 1 Aquilian Liability* (1984) 1; Burchell *Principles of Delict* (1993) 3-4; Hutchison "The Nature and Basis of Contract" in *The Law of Contract in South Africa* (eds Hutchison & Pretorius) (2012) 8-9.

39 Hutchison 8. See also Lubbe & Murray *Farlam & Hathaway: Contract* (1988) 1: "Contractual obligations are created by agreement (or apparent agreement) of parties. Unlike many other obligations, they are supposed to arise voluntarily. An obligation based on delict, on the other hand, arises *ex lege* when a legal subject wrongfully and without adequate justification,

Thus, in the law of delict the equality postulate is animated imperatively, when the plaintiff demands satisfaction by way of litigation. Not unlike the criminal law, then, the law of delict also operates in coercive conditions, with the defendant being drafted against his will into a determinate legal relation with the plaintiff, to answer for the damage which his errantry has occasioned. His conduct has violated the principle of equivalence and introduced disproportion into his relationship with the plaintiff. The law operates to countermand the inequality between the parties caused by the delict. Contractants have an *ex ante* commitment to the principle of equivalence; with a delict, the principle of equivalence is deployed *ex post facto* at the instance of the plaintiff.[40]

The overlap between the law of delict and criminal law around the principle of equivalence may be extrapolated directly from the routine coincidence of circumstance between delict and crime. According to Van den Heever:

> A delict is not a distinct factual concept; it is merely a wrong regarded from the individual's point of view and in the light of procedure. When the state assumes the right to pursue a wrong, to exact punishment and so effect atonement we call the proceedings criminal and the wrong, regarded from this point of view, a crime. The state may also allow the individual directly harmed by the wrong to sue for a readjustment of his interests infringed by it.[41]

Thus, delict and crime occupy more or less the same factual space insofar as "all of the crimes against person and property are also delicts".[42] The distinctions between them are socio-legal and pertain primarily to their definitional limits and juridical classification. As a result of the different jurisprudential lenses through which they are viewed, crime is understood to have an eminently public law face whereas delict is demarcated in private law terms. However, the disparate perspectives cannot obfuscate completely their common contractual countenances. As seen above, the contractualisation of delict to enforce the principle of equivalence is uncontroversial. There appears to be no opposition of substance to the idea that the plaintiff's demand for recompense according to the contractual principle of equivalence is dependent upon the *ex lege* and *ex post facto* imposition of delictual obligations upon the defendant. Since delict and crime are juridical obverses, it seems logical to extend the contractual view of delict to crime. In this context, the Pashukanist formulation of crime as an involuntary contract concluded *ex post facto* acquires a meaning which is both sensible and defensible. In an economy of generalised commodity production, the principle of

39 intentionally or negligently, infringes a recognised interest of another to the detriment of that person."
40 See Loubser "Concurrence of Contract and Delict" 1997 *Stell LR* 113-114.
41 Van den Heever *Aquilian Damages in South African Law: Volume 1* (1944) 2. See also Turner *Kenny's Outlines of Criminal Law* (1962) 1.
42 Burchell *Principles of Delict* (1993) 2. See also Snyman 4; Boberg 1-3.

equivalence is the necessary fundament of the legal form and juridical relations needs to be premised upon subject equality. Both crime and delict obey this juridical requisite, albeit from opposite ends of the public law-private law divide.

The intersections of contract, delict and crime reflect the fact that their differences are more ostensible than intrinsic, given that one course of misconduct may qualify as either a breach of contract or as a delict, and that behaviour which is *prima facie* criminal simultaneously may be delictual. Contract, delict and crime evidently comprise an integrated circuit of liability structured by a collective obeisance to the principle of equivalence. Burchell observes that crime and delict differ formally but not materially.[43] This insight may be extended readily to the opposition between contract and delict, to produce a triad of commingled concepts each substantively indistinguishable from the others. From a Pashukanist perspective, however, it is the contract component of the triad which is pivotal. Contract may be comprehended as the *grundnorm* from which is derived the criminal and delictual components. In other words, crime and delict are variants of the equality postulate which is the constitutional core of contract.

Hence, the homology between the commodity form and the legal form is valid also for criminal law. Indeed, Pashukanis considers that "the characterisation 'criminal law' becomes utterly meaningless if this principle of the equivalent relation disappears from it".[44]

Such is the intended reach of the general theory: The notion of equivalence which is a structural feature of every contract is comprehended also to be a compositional trait of every crime.[45] His critics notwithstanding, Pashukanis's approach to the criminal law is neither crude nor contrived.

3 Punishment and the Principle of Equivalence

Pashukanis embraces the equality postulate as a theoretical imperative. Unsurprisingly, therefore, he invokes it also in his analysis of punishment. For him, punishment is about equivalent requital. It is an exchange transaction which places the offender on an equal footing with his victim. The offender has asserted a claim to unrequited priority over his victim; punishment puts him in his place, literally, by coercing his

43 Burchell *Principles of Delict* (1993) 57.
44 Pashukanis 176.
45 See Arthur "Editor's Introduction" in Pashukanis *Law and Marxism: A General Theory* (1978) 15; Jakubowski *Ideology and Superstructure in Historical Materialism* (1990) 49. See also Stone "The Place of Law in the Marxian Structure-Superstructure Archetype" 1985 *Law and Soc R* 44-45. Interestingly, Stone, who is no friend of Pashukanis, considers that Pashukanis's "views on criminal law are insightful" and submits that they "may be analogised to many other areas of law", identifying "tortious conduct" as one such area.

acquiescence in the principle of equivalence.[46] Punishment, like crime, is governed by the law of the commodity. The criminal sanction is the performance due by the offender under the contract which perforce he has concluded with his victim. If crime is the violation of the equality postulate then punishment is its vindication.

Pashukanis's theory of punishment has elicited objections similar to those heaped upon his theory of crime. For example, Hunt alleges vociferously that Pashukanis's theorisation of a correspondence between the commodity form and punishment is tantamount to lexical sleight of hand:

> The weakness of his treatment lies precisely in the fact that the identity he seeks to establish lies in nothing more than the verbal equation achieved by the dual usage of equivalence and the assertion that the verbal correspondence evidences a real correspondence.[47]

For Hunt, it seems, Pashukanis is a latter-day Schoolman stretching the subtleties of logic beyond their legitimate limits. Warrington, too, will have no truck with the theory of equivalent punishment and dismisses it summarily and contumeliously as "faintly comic".[48] The remainder of this section attempts to defend Pashukanis against such criticism.

As with his theory of crime, Pashukanis's theory of punishment is grounded in very reputable historical precedent, this time in the jurisprudence of Hugo Grotius. The Dutch master was unequivocal about punishment as equivalent requital: "It is undoubtedly one of the first principles of justice to establish an equality between the penalty and the offence".[49]

His explanation of this first principle of justice expressly attributes an unmistakable contractual texture to the nature of punishment:

> Now in the eye of the law, every penalty is considered as a debt out of a crime, and which the offender is bound to pay to the aggrieved party. And in this there is something approaching to the nature of contracts. For as a seller, though no *express* stipulation be made, is understood to have bound himself by all the *usual*, and *necessary* conditions of a sale, so, punishment being a natural consequence of crime, every heinous offender appears to have *voluntarily* incurred the penalties of law.[50]

Grotius highlights punishment as a voluntary aspect of the contract between offender and victim. This voluntariness may be understood as supplementary to the involuntariness of the contract as a whole. In other words, if crime entails the retrospective imposition of a contract upon the offender, then that contract entails a prior conscious acceptance by the

46 See Ripstein 160-161.
47 Hunt *Explorations in Law and Society: Toward a Constitutive Theory of Law* (1993) 81.
48 Warrington 62.
49 Grotius (trans A Campbell) *The Rights of War and Peace* (1979) 2 20 2.
50 Grotius 2 20 2, original emphasis.

offender of the punishment prescribed by law for his crime. In this connection, every decision to break the law may be taken as "a voluntary contract to submit to punishment".[51] Grotius considered this conception of punishment to be Aristotelian in origin:

> For, as Michael the Ephesian observes on the fifth book of Aristotle's *Nicomachean Ethics*, the ancients gave the name contract, not only to the voluntary agreements which men made with each other, but to the obligations arising from the sentence of the law.[52]

All in all, then, it seems that Pashukanis took his cue from jurisprudential sources which confer upon the contractual theory of punishment the same historical legitimacy accorded the contractual theory of crime. In the Pashukanist analytic, crime and punishment constitute a juridical dialectic which is mediated by the principle of equivalence embedded in their common contractual constitution.

Historically, the evolution of the principle of equivalence echoed the evolution of the commodity economy. However, commodity production emerged comparatively late in the pantheon of social development. Humankind's aboriginal mode of production was primitive communism, a form of social organisation born of the precariousness of existence in inhospitable natural conditions. The fundamental prehistoric social unit was the primitive commune which was premised upon the practice of equality in all things.[53] Prehistoric justice was blood revenge: given the ethos of egalitarianism, an injury to any member of the primitive commune, because it put the existence of the commune at risk, was experienced as an injury to the commune as a whole.[54] Blood revenge was a mechanism of self-preservation[55] and meant that any and all members of the victim commune could exact revenge against any and all members of the offender commune.[56] Thus, blood revenge was indiscriminate, and would lead routinely to the blood feud, that is, an open and inconclusive vendetta between communes.

However, over time the primitive communards came to appreciate the real dangers of mutual destruction embedded in the blood feud, and transformed it into the *lex talionis*. The law of retaliation determined that revenge had to be moderated by the principle of equivalent requital. In other words, the blood feud had to be controlled to the extent that there had to be a literal correspondence between harm suffered and revenge taken: an eye for an eye, a tooth for a tooth and so forth.[57] The prehistoric *lex talionis* meant, essentially, the customary regulation of blood revenge according to the primitive communist commitment to

51 Grotius 2 20 2.
52 *Ibid*.
53 See Reed *Woman's Evolution: From Matriarchal Clan to Patriarchal Family* (1975) 197-198; Manfred *A Short History of the World: Volume 1* (1974) 12.
54 See Briffault *The Mothers* (1959) 267.
55 See Lafargue *The Evolution of Property* (1975) 161.
56 See Lafargue 166.
57 Pashukanis 168.

complete equality. During this epoch, commodity production was non-existent for the most part and episodic only during the transition from prehistory to history. Hence, the original *lex talionis* did not bear the imprint of the commodity form and its equality postulate.

Civilisation commenced with the break-up of the primitive commune and the installation of the commodity form as the economic centrepiece of the new society.[58] The penal correlate of this historic socio-economic transformation was the shift from blood revenge to composition.[59] The latter was "a system of expiatory payment"[60] which allowed offenders to pay for their infractions, literally, provided they could afford to do so.[61] It transformed the talionic notion of an eye for an eye into the pecuniary notion of an eye for value.[62] As Ripstein puts it:

> Historically, criminal law begins to supplant revenge when the ideas of compensation and finality are introduced. The idea of compensation is always a commodity-bound idea, because it is the idea of the equivalent.[63]

Composition was the *lex talionis* commodified. With its advent, equivalence took on an economic character. Literal equivalence gave way to economic equivalence, and biological formulae were replaced by calculation of abstract values.[64] The juridical moment had arrived.

For an institution to be juridical it must adhere to the principle of equivalence.[65] Punishment, in its evolutionary aspect, is such an institution. Its genesis coincides with the genesis of the commodity form and its progression mirrors the advance of commodity production and circulation. It is constituted in terms of equivalent requital: the offender has to pay for the harm he has caused, and the payment must be commensurate with the degree of harm suffered by the victim.[66] In this context, the criminal sanction becomes "a form of exchange, a peculiar form of circulation, which has its place alongside 'normal' commercial circulation".[67] For Pashukanis, then, punishment is a type of commodity exchange.

---

58 See Mandel *Introduction to Marxism* (1979) 19; Novack *Understanding History: Marxist Essays* (1980) 44-45.
59 See Newman *Law and Economic Organisation* (1983) 163.
60 Pashukanis 168.
61 The indigent were excluded from the system of composition *ab initio* and had to endure the pain and indignities of physical punishment for their transgressions.
62 See Lafargue 174.
63 Ripstein 251.
64 Pashukanis 170. See also Melossi & Pavarini 2-3: "The transition from private vendetta to retributive punishment, that is, the transition from an almost 'biological' phenomenon to a juridical category, requires as a necessary precondition the cultural dominance of the concept of equivalents based on exchange value."
65 Pashukanis 170 theorises an identity between "the juridical idea" and "the idea of the equivalent".
66 See Pashukanis 169.
67 Pashukanis 176.

As an attribute of the commodity economy, punishment is governed by the measure of value of every commodity: labour and time. The form of punishment in which the development of criminal justice under capitalism has summated is imprisonment, that is, the exchange of a determinate portion of the offender's freedom, measured in time, for the harm his crime has caused. As Pashukanis demonstrates, the jail term is an obvious derivative of the exchange transaction which underlies the legal form. Despite the emergence of many other forms of criminal sanction, incarceration has remained dominant in all capitalist social formations. It has turned out to be the hardy perennial.[68]

Imprisonment articulates the principle of equivalence most completely, more so even than the pecuniary sanction. Not every offender can afford to pay a fine. But every offender can be incarcerated. The prison (despite whatever sordid conditions the offender may have to endure) is the ironic flagship of equality in the criminal justice system. Imprisonment is the penal materialisation of the principle of equivalence. It is the paradigmatic means whereby the state is able to recover the juridical relation which the crime has infracted, and to secure the preservation of the legal form.

Imprisonment, in any form, including periodical imprisonment and imprisonment with hard labour,[69] is essentially an exchange transaction, in which the currency is freedom, measured in determinate time periods. For Pashukanis:

> Deprivation of freedom, for a period stipulated in the court sentence, is the specific form in which modern, that is to say bourgeois-capitalist, criminal law embodies the principle of equivalent recompense.[70]

He continues:

> The offender answers for his offence with his freedom, in fact with a portion of his freedom corresponding to the gravity of his action. This conception of liability would be quite superfluous in a situation where punishment has lost the character of an equivalent. Were there really no trace of the principle of equivalence remaining, then punishment would entirely cease to be punishment in the juridical sense of the word.[71]

It is a logical requisite of the commodity economy that its penal regime be centred upon the prison sentence. It is the great leveller, not only as

68 See Melossi & Pavarini 185, original emphasis: "Punishment in prison – as the deprivation of a *quantum* of liberty – becomes punishment *par excellence* in a society producing commodities; *the idea of retribution by equivalent* thus finds in prison punishment its most complete realisation precisely in so far as (temporary) loss of liberty represents the most simple and absolute form of 'exchange value'".

69 Hudson "Punishment and Control" in *The Oxford Handbook of Criminology* (eds Maguire *et al*) (2002) 251 notes "the return of hard labour" in contemporary imprisonment regimes.

70 Pashukanis 180-181.

71 Pashukanis 179.

regards the inequality between offender and victim, but also as regards offenders *inter se*.

Incarceration creates the abstract offender, the generic criminal, who concentrates in his person all manner and method of criminal behaviour. It is thus a device of equalisation. It ensures that the offender gives satisfaction to the victim, and that he does so on terms which are equal relative to every other offender. The expanded reproduction of the commodity economy depends crucially upon the ordered reproduction of the legal form. The prison sentence is the one form of punishment which has proved itself indispensable thus far to the reproduction of the legal form in the sphere of criminal justice.

Punishment, then, is necessarily about equivalent requital. It is a conception which is derived historically from the dissolution of the natural economy and the concomitant evolution of generalised commodity production as an economy of labour time. Pashukanis explains:

> For it to be possible for the idea to emerge that one could make recompense for an offence with a piece of abstract freedom determined in advance, it was necessary for all concrete forms of social wealth to be reduced to the most abstract and simple form, to human labour measured in time.[72]

The archetypal capitalist form of criminal punishment thus is linked intimately to the archetypal capitalist form of production and exchange. The spirit of the commodity is so ubiquitous that it penetrates even the steel gates and concrete walls of the prison.

As the penal manifestation of commodity exchange, the notion of equivalent requital applies also to all other forms of punishment, including such non-custodial sanctions as the fine, correctional supervision, property forfeiture and community service. They, too, are structured according to the desideratum of parity which makes of state punishment the juridical phenomenon that it is. Pashukanis observes:

> In principle, punishment in keeping with guilt represents the same form as retaliation in proportion to the injury. Its most characteristic feature is the arithmetical expression of the severity of the sentence: so and so many days, weeks, and so forth, deprivation of freedom, so and so high a fine, loss of these or those rights.[73]

The form of the criminal sanction is thus of little consequence. The essence of each form, whether custodial or non-custodial, is given by the principle of equivalent requital. As Pashukanis notes, this principle is an

72 Pashukanis 181. See also Melossi & Pavarini 184-185: "The idea of the deprivation of an abstractly determined quantity of liberty, as the dominant form of penal sanction can in fact only be realised with the advent of the capitalist system of production, that is, in that economic process which reduces all forms of social wealth to that most simple and abstract form of human labour measured in time".

73 Pashukanis 180.

"essentially absurd idea". From the non-juridical point of view it is nonsensical to attempt to equate every harm occasioned by every crime with deprivation of freedom measured in time. But this is an absurdity which cannot be avoided for "so long as the commodity form and the resultant legal form continue to make their mark on society".[74] A rational basis for punishment is possible only if and when it may be reconstructed outside of the juridical paradigm. And it likely will take a social revolution to transcend this paradigm and deprive it of its current authoritative status.

In sum, Pashukanis's theory of punishment accords completely with the commodity form theory of law. Certainly, there is nothing either comical or concocted about comprehending the criminal sanction as a variant of commodity exchange. Warrington deployed mockery as a weapon of criticism without making any attempt whatsoever to ground it analytically.[75] Hunt sought to trivialise Pashukanis's theoretical sophistication by portraying it as a scholastic rhetorical device. Such facile attempts at critique matter little for a theory of punishment which is anchored in the materiality of the commodity form.

4 Crime, Class and Coercion

The state is a party to every criminal justice transaction. The offence is construed as a contravention of a state norm and the state generally directs the prosecutorial process. Criminal justice is, in a word, an affair of state. But the contemporary state is the capitalist state, and criminal justice is thus also capitalist justice. This is the context of Pashukanis's contention that, in the capitalist epoch, state hegemony over the criminal justice system is one of the weapons at the disposal of the bourgeoisie to protect its class rule and to fend off the demands of the dominated classes.[76] Indeed, Pashukanis tells us that "criminal justice in the bourgeois state is organised class terror".[77] Such is the class content of the criminal law. It is that branch of the law which expresses most directly the violence immanent in the rule of the bourgeoisie, including its most advanced legal form of the *Rechtsstaat*.[78]

74 Pashukanis 185.

75 Norrie "Pashukanis and the Commodity Form Theory: A Reply to Warrington" 1982 *International Journal of the Sociology of Law* 432 makes the following astute observation about Warrington's effort to ridicule Pashukanis's theory of punishment: "This amusing quality of the theory is apparently self-evident since we are not told why it is faintly comic."

76 Pashukanis 173. It must be noted that the class nature of the state is not Pashukanis's primary concern. He deals with it only after he has established the role of the state in the reproduction of the commodity and legal forms. These two aspects of "bourgeois statedom" operate at distinct levels of abstraction, the latter at a higher level than the former. It is the latter aspect which is integral to the general theory of law.

77 Pashukanis 173.

78 See Jessop *The Capitalist State: Marxist Theories and Methods* (1982) 84-85.

Hunt considers that Pashukanis's highlighting the class content of the criminal law imports a dualism into his analysis of law:

> Thus he introduces a sharp polarity between two modes of law, the criminal law as a means of securing class domination and the civil law as the mechanism governing the exchange relations between atomised legal subjects.[79]

Hirst, too, is concerned about the disintegrative impact of class conflict upon the regular contours of the legal form:

> For Pashukanis criminal law is merely the derivative extension of the form of law (a form which has a necessary and autonomous function as private law, the regulation of production through property) to class oppression. In conditions of acute class conflict the ideological nature of the legal form in criminal law is revealed, the juridic mask is discarded to reveal class force unlimited by legal rules or procedure.[80]

Thus, both Hunt and Hirst perceive a deep divide in the Pashukanist general theory, which purportedly bifurcates law into criminal law and private or civil law. And whereas civil law is consonant with the principle of equivalence embedded in the commodity form, criminal law merely pays lip service to this principle, exploding it regularly in favour of the coercion of class command and control. The implication of such an alleged dualism is, of course, to endorse the argument that the fit between the commodity form theory of law and the criminal law is, at best, an awkward one. Indeed, it appears, from Hunt's and Hirst's perspective, that the criminal law is not subsumed within the homology between the legal form and the commodity form and, hence, that the so-called general theory, after all, is pretty circumscribed in its juridical ambit and explanatory power.

However, this argument fails to comprehend the theoretical requisite that the role of the state as an instrument of class control has to be assessed on a quite different level of abstraction from its relation to the general theory of law. To be sure, the class analysis of the role of the state ought not to be conflated with the analysis of the state from a juridical perspective. The former is essentially a political analysis and is concerned to expose the day-to-day complicity of the state in the dictatorship of the bourgeoisie; the latter operates at a higher level of abstraction and is concerned to make sense of the role of the state in relation to that juridical "absurdity" denominated in the principle of equivalence. Certainly, the fact that the criminal law is Janus-faced, secreting behind the legal form its connivance in bourgeois class terror does not place it outside the pale of nor does it bifurcate the general theory. The real issue is to make sense of the role of the criminal law, as a state competency, in relation to the legal form and its private law manifestations. Jakubowski's insights are especially valuable here:

79 Hunt 81.
80 Hirst "Introduction" in Edelman *Ownership of the Image: Elements for a Marxist Theory of Law* (1979) 10.

> The legal form and the commodity form work their way into the existing relations hand in hand. As a relatively independent political power develops, serving the interests of the ruling class as a whole, so too does public law. Public law regulates the relations between the state and public institutions, and between these and the citizens; it serves to execute and protect private or civil law by means of the power of the state. The foundation of all these relations is still legal subjectivity and the recognition of the legal capacity of man, which give the relations of domination a general form.[81]

As a branch of public law, criminal law evidently helps to create the juridical milieu required for the uniform operation of contract law, property law, the law of delict and the like. It simultaneously expresses and facilitates the workings of the legal form. The criminal law was not imbricated in the birth of the legal form, but its imprimatur is indispensable to the continued existence and reproduction of that form.

Criminal law may not be linked as conspicuously to the principle of equivalence as the law of contract and the other branches of private law. Yet, it is the branch of law in which the legal form is most conspicuously present, for it is here that the legal subject is found in its most impersonal and abstract form. It is in the criminal trial that "the juridical element first and most crudely detaches itself from everyday life and becomes fully autonomous".[82] It is here that the juridical moment peaks, in the "transformation of the actions of a concrete person into the proceedings of a legal party, that is of a legal subject".[83] In other words, criminal law sets the high-water mark for legal intercourse. It is the branch of law which depersonalises actors most fully in order to equalise them as legal subjects most completely. Criminal law is synecdochic for law itself.[84] It represents, in telescoped terms, all the characteristics of the legal form. There is no more loyal commitment to the notion of equivalence than the literal replacement of the victim by the state in a criminal matter. Prior to the replacement, the offender, as individual, had lorded it over the victim, as individual, and unilaterally had violated the principle of equivalence. With the intervention of the state, both offender and victim are transformed from specific individuals into general legal subjects, the advantage of the offender is eradicated, and the principle of equivalence is rehabilitated.

The coercive constitution of criminal law is the guarantor of the *Rechtsstaat* as exemplar of "market relations among formally free and equal individuals".[85] From this perspective, criminal law is as much a part of the legal form as any other branch of law. Criminal law is not ejected from the compass of the legal form because of its terroristic aspect; on the contrary, it cements their interdependence. Criminal law is a second-generation materialisation of the legal form and serves to

81 Jakubowski 49.
82 Pashukanis 167.
83 *Ibid*.
84 *Idem* 168.
85 Jessop 85.

ensure the propagation of both the legal form itself and its first-generation manifestations. Critics such as Hunt and Hirst have been quick to perceive a contradiction between criminal law and the legal form and hence to allege that the commodity form theory does not withstand scrutiny as a general theory. However, scrutiny of the critics' concerns reveals quickly that they are groundless. For example, Hunt argues that Pashukanis commits theoretical heresy trying to reconcile private law and criminal law:

> To provide a bridge that overcame the dualism between private law and state law he introduces what is undoubtedly the weakest feature of his general theory. He subsumes his attempt to theorise state and criminal law into a theory of punishment.[86]

This is a proposition which beggars belief. Pashukanis's theory of punishment is just that, a theory of punishment derived from his theory of law. There is nothing in his work to suggest that he was attempting the converse, to extrapolate his theory of punishment into a theory of law. Also, Pashukanis formulated a theory of crime which is by no means a by-product of his theory of punishment and which was constructed directly from the fundamentals of the general theory. Hunt wants to have his cake and eat it: he charges Pashukanis with dualism; and then charges him with theoretical aberrance in attempting to resolve the dualism. However, both charges are trumped up, with neither having a basis in anything which Pashukanis postulates. And certainly, Pashukanis did not conflate levels of abstraction as do his critics when they attempt to fragment the general theory of law or even to annihilate it.

5 The Public Law-ness of Criminal Justice

Criminal law is a branch of public law and criminal justice is statist. The entire criminal justice process takes place under the auspices of the state and its juridical apparatuses. Against this backdrop, Stone alleges: "Pashukanis's approach does not explain why the state, and not the victim or his or he relatives, is the plaintiff in criminal cases."[87]

Pashukanis, it appears, is unable to account for the public nature of criminal law. In other words, Stone reckons that the commodity form theory of law is deficient for neglecting to elucidate the statist structure of criminal justice.

---

86 Hunt 81.

87 Stone 45. Ignatieff "State, Civil Society and Total Institutions: A Critique of Recent Social Histories of Punishment" in Cohen & Scull (eds) *Social Control and the State: Historical and Comparative Essays* (1985) 96-99 has raised the same issue by challenging the "assumption in Marxist social theory" that capitalism requires the "state penal sanction", and suggesting that the threat of force is not necessary for the reproduction of "exploitative social relations".

Stone's concern is misplaced. Pashukanis comprehended fully that criminal law was coercive in its operations. However, he understood also that the fundamentally juridical nature of the commodity economy requires such coercion to be public, in the sense that it is separated formally from the exercise of personal power. Thus, he observes:

> Coercion as the imperative addressed by one person to another, and backed up by force, contradicts the fundamental precondition for dealings between the owners of commodities … For in the society based on commodity production, subjection to one person, as a concrete individual, implies subjection to an arbitrary force, since it is the same thing, for this society, as the subjection of one owner of commodities to another.[88]

The core principle of equivalence which defines the commodity economy is violated, and the reproduction of the commodity economy itself is rendered precarious, if the coercion which structures it is personalised or privatised. Pashukanis again states:

> This is also why coercion cannot appear here in undisguised form as a simple act of expediency. It has to appear rather as coercion emanating from an abstract collective person, exercised not in the interest of the individual from whom it emanates … but in the interest of all parties to legal transactions.[89]

In other words, a public authority with coercive competence is inscribed in the constitution of the commodity economy. So too, by extension, is the public nature of criminal justice, as the branch of law which is concerned most, and most conspicuously, with the exercise of that coercive competence.

The Pashukanist position reduces to the proposition that capitalism needs a public criminal justice system. That is the answer to Stone's complaint.[90] Private law may be the "natural" law of the commodity economy, and may provide the foundation of all legal discourse and interaction. But it is public law, and criminal law in particular, which is necessary to secure the conditions for the reproduction of the juridical physiognomy of capitalism. It is not possible for the bourgeoisie to entrust its criminal justice system to the operations of private law, as it has its contractual and proprietary regimes. In other words, criminal justice in the commodity economy cannot but be statist in its essentials.[91]

---

88 Pashukanis 143.

89 *Ibid*.

90 It is also the answer to Ignatieff's challenge to the necessarily statist texture of criminal justice.

91 There has been a recent rapid growth of private prisons. See Bates "Prisons for Profit" in Haas & Alpert *The Dilemmas of Corrections: Contemporary Readings* (1999) 595; Morgan "Imprisonment: A Brief History, the Contemporary Scene, and Likely Prospects" in *The Oxford Handbook of Criminology* (eds Maguire *et al*) (2002) 1147-1149. However, this development does not entail a move away from the basic idea of state punishment. The private prison continues to enforce a public disposition of criminal matters.

Justice in the *Rechtsstaat* must proceed from the equality postulate. The parties to a legal transaction may not be equal in fact and one may have at his disposal many more resources than the other. But the stronger party cannot expect to rely upon this power differential being translated into an automatic legal advantage. Formally, at least, he has to accept the equality postulate which makes his counterpart his peer at law.[92] This is the necessary consequence of the uniform legal subjectivity conferred upon them by the juridical complexion of the commodity economy.

However, every crime plunges the idea of legal subjectivity and its accompanying equality postulate into crisis. Every crime is an exercise in inequality, a practical failure of the principle of equivalence. It is a manifesto of personal power or "private rationality" against the juridical and its core value of equivalence. The offender asserts his interests to be prior to those of his fellow legal subjects, and thereby approves "the subjection of one owner of commodities to another".[93] His crime is a revolt against the levelling effect of legal subjectivity in the commodity economy. He is a true champion of "the concrete individual" against its negation, the abstract legal subject,[94] and he is not averse to one such concrete individual enthralling another. He is a recalcitrant legal subject, challenging the logic of the juridical worldview which is integral to the ideational life of the capitalist mode of production.[95] He is an inveterate anti-egalitarian, a devotee of the pre-legal, who lives surely, if unconsciously, by Marx's famous dictum that equal right is, in its content, a right of inequality.[96]

By committing the crime, the offender has adjudged the constraints upon arbitrariness inscribed in the sphere of privatised legal relations to be inadequate to the defence of and, by the same token, the re-assertion of the equality postulate. The fate of the latter thus becomes a public concern, that is, a matter for state action. Logically, that which the offender has sundered cannot be relied upon to reconstruct him as a compliant legal subject. Criminal justice is thus inexorably public justice. Only the state, as the embodiment of the acquiescent legal subject, has the capacity to extract from the offender the conformity which is required to reconcile him to the idea of legal subjectivity, both his own and everybody else's. It is one of the dialectical ironies of the legal form that the efficient reproduction of the private sphere is dependent, ultimately and unavoidably, upon the public power. The public is the true guarantor of the private.

92 See Pashukanis 143-144.
93 *Idem* 143.
94 *Ibid*.
95 See Engels "Lawyers' Socialism" in Marx & Engels *Collected Works: Volume 26* (1990) 598 who describes the juridical worldview as "the classical one of the bourgeoisie".
96 Marx "Critique of the Gotha Programme" in *The Marx-Engels Reader* (1978) 530 (ed Tucker). The offender likely is blissfully unaware of both the man and his dictum.

As noted, criminal justice is perhaps the most eminently public branch of law. It is in this sphere where the violent nature of the law is most evident, and hence it is here where it is necessary to restrict the use of coercion to the formal organs of the state.[97] The private settlement of criminal disputes is not properly legal, in that it is not an organ of state which has the final say in the disposition of the matter. Such private resolutions either hark back to a pre-legal past or portend a post-legal future. Criminal justice is thoroughly legal. Not only is it is dispensed by the state, but it is also the only variant of justice which involves the state as a party. It cannot be otherwise.

Civil and criminal cases differ essentially in the fact that the former involve private parties on both sides while the latter involve a public party on one side. On the face of it, both parties to a civil matter proceed from the premise that the principle of equivalence is a valid one. Of course, each is operating from self-interest, submitting that the other has violated the principle, and both seek from the court a decision that will re-establish the principle *inter se*. While the submissions of either are entirely self-serving, there is no argument about the cogency of the juridical relation. Civil disputes are about using that relation to self-advantage. They do not require state intervention, except as ultimate arbiter regarding the interpretation of the principle of equivalence.

The offender has no such agenda. He does not seek to use the principle of equivalence to his advantage. Instead, he rejects the principle, albeit usually unthinkingly or unwittingly. He asserts a claim to operate outside the parameters of law, and not only in the sense of breaking the letter of the law. Objectively, a crime is, therefore, an act of outright subversion of the legal relation. The offender has no stake in co-operating or even competing with the victim in curial proceedings to settle the interpretation of the principle of equivalence. In these circumstances, it is necessary for the state to become a party to the matter and to compel the offender to respect the principle and to participate in its operation. If a crime is a repudiation of the principle of equivalence, then the criminal trial is its affirmation.

It has been held that: "From a Marxist perspective, state control over deviance is an integral feature of modern capitalism and is not likely to be relinquished."[98] The truth is that relinquishment of "state control over deviance" is not only unlikely but also unthinkable. Such control is an indispensable feature of modern capitalism and its criminal justice system and hence cannot be forsaken. The capitalist social formation requires a criminal justice system which operates under the aegis of its state. The notion of a private criminal justice does violence to the idea of the legal subject who lives (and dies) by the principle of equivalence.

97 Seagle *The History of Law* (1946) 6-7.
98 Minor & Morrison "A Theoretical Study and Critique of Restorative Justice" in *Restorative Justice: International Perspectives* (1996) 124.

Despite their origins in private law, the legal form and its concomitants cannot survive without the patronage of the state.

6 Excursus: Pashukanis in South Africa

There is a vibrant and admirable jurisprudence promoting constitutional democracy and defending human rights in South Africa. However, the commodity form theory of law falls outside the ambit of this jurisprudential matrix and seldom, if ever, is given considered philosophical attention. This neglect of Pashukanis may be due to the fact that South African jurisprudence, in the main, yet has to confront the capitalist character of the South African social formation. The capitalist mode of production long has been dominant in South Africa, and certainly throughout both the *apartheid* and the post-*apartheid* eras. Whereas *apartheid* may have obscured somewhat the capitalist fundamentals of society, the transformation to a non-racial democracy has rendered them starkly visible. Indeed, this politico-legal transformation was simultaneously a process of normalisation: South Africa lost its racist exceptionalism and took its place as a regular member of the capitalist world system.[99]

Capitalist normalisation calls for a jurisprudence which transcends the conventional liberal concern with liberty, equality and the rule of law. Liberal jurisprudence stands perplexed in the face of the scandalous socio-economic inequalities and the epidemic of criminality, violent and economic, which continue to embarrass a country in the van of constitutional democracy internationally. These debilities are linked intimately to its idealist conception of law as a system of norms, obstructing comprehension of the material foundations of legal relations.

Pashukanism is about excavating the material basis of the juridical in the commodity economy and apprehending the structural comity between legal relations and capitalist relations of production. It has been argued persuasively that "without a theory of the legal form, the specificity of law itself is impenetrable".[100] Pashukanism supplies the theory of the legal form which liberal jurisprudence lacks. The commodity form theory of law well could provide the analytical roadmap needed to comprehend the explosive contradictions between the politico-legal and the socio-economic which bedevil the South African transformation, and to grasp the logic of the post-*apartheid* crime emergency. And it well could hold the key to the development of a jurisprudence which engages properly the capitalist context of legal relations. Certainly, a South African jurisprudence which continues to ignore Pashukanis is depriving itself of a rich and robust epistemological resource.

---

99 See generally Alexander *An Ordinary Country: Issues in the Transition from Apartheid to Democracy in South Africa* (2002).

100 Miéville *Between Equal Rights: A Marxist Theory of International Law* (2005) 2.

7 Conclusion

The private law origins of the commodity form theory of law cannot constitute an ontological barrier to its extension to public law. Pashukanis's derivation of the general theory from the law of contract was a necessary one, dictated by the real history of the legal form. Of course, theories cannot be confined within the narrow limits of their birth, and the generalisation of the commodity form theory was a natural occurrence, more or less. What is more, unlike so many other general theories of law which tend to be ahistorical, Pashukanis's has the virtue of laying bare its own material parentage in the evolution of the commodity economy.

This article has rebutted the argument that the general theory is insufficient in that it is unable to explain crime and its punishment. It has shown that substantial aspects of the criminal justice process and of criminal law may be read as manifestations of the principle of equivalence. It has demonstrated, also, that the principle of equivalence is able to account for criminal punishment easily and without resort to rhetorical device. Of course, the general theory cannot explain all the specificities of criminal justice doctrine. However, that is not its function. No general theory of law can be expected to traverse every nook and cranny of every branch of law. The primary task of the Pashukanist general theory of law is to clarify why law as a whole takes the form it does. It achieves this task with some considerable proficiency and élan. Thus, it may be contended confidently that there exist "good grounds for taking seriously Pashukanis's claim that his is a general theory of law, and not one illicitly generalised from private civil law".[101] In its turn, South African jurisprudence would do well to take Pashukanism seriously as a general theory of law.

101 Norrie 434.

Surrogate motherhood agreements and their confirmation: A new challenge for practitioners?

Caroline Nicholson
BProc LLB LLM LLD
Professor, Department of Jurisprudence,
University of Pretoria

Andrea Bauling
BA LLB
Lecturer, Department of Jurisprudence
University of South Africa

OPSOMMING

Surrogaatmoederskap-ooreenkomste en hul Bevestiging: 'n Nuwe Uitdaging vir Praktyns?

Ondanks die feit dat onlangse regspraak sommige onduidelikhede aangaande die regsposisie van surrogaatskap uit die weg geruim het, bestaan verskeie regs- en etiese dilemmas steeds op dié gebied. Ten spyte van die huidige *lacunae* dien die kontrak as 'n hulpmiddel om die belange van beide die lasgewende ouer(s) en die surrogaatmoeder te bevestig en te beskerm.

Die howe vereis 'n volledige, eerlike en omvattende kontrak, beëdigde verklaring en aansoek. Ten spyte hiervan blyk partye onbewus of ongeërg oor die moontlikheid en uitwerking van hul versuim om gehoor te gee aan hierdie vereistes. Voordat surrogaatskap deur die Kinderwet[1] gereguleer was, het onvrugbare paartjies waarskynlik surrogaatskap onwettiglik beoefen. Die vraag is dus waarom, nou dat die praktyk wettig is, paartjies steeds die gereg tart?

Vanweë die aard van surrogaatskap bestaan die praktyk grotendeels uit die strewe na 'n balans tussen die belange van die verskeie partye, die kind en die gemeenskap as 'n geheel – 'n taak wat byna onmoontlik blyk te wees. Tans is die surrogaatmoederskapsooreenkoms die mees gepaste instrument om die belange van alle betrokkenes te beskerm. Die moontlikheid van 'n surrogaatskap-spesifieke model moet ondersoek word, om sodoende voornemende partye, asook ons howe leiding te bied in die uitleg, bekragtiging en implementering van surrogaatmoederskapsooreenkomste onder Hoofstuk 19 van die Kinderwet.

Todat 'n surrogaatskap-spesifieke model ontwerp word, of 'n hof 'n uitspraak lewer wat só 'n model impliseer, word daar aanbeveel dat die kontraktereg-model toegepas word. Dit maak voorsiening vir die moontlikheid dat die bedoeling van die partye nagejaag kan word. Verder word ook aan die hand gedoen dat toepassing van die kontraktereg-model

1 'n Nie-amptelike vertaling van die Children's Act 38 of 2005 is beskikbaar by http://www.vra.co.za/vorms/KINDERWET%20(No%2038%20van%202005)%20Afr.pdf.

kan lei tot die teëwerking van die onsekerhede wat daargestel word deur die afwesigheid van regulasies kragtens Hoofstuk 19 van die Wet.

1 Introduction

This article will examine surrogacy law in South Africa with a view to making some suggestions for the implementation of regulations to chapter 19 of the Children's Act.[2] Surrogacy remains a controversial topic both in South Africa and elsewhere. There has been considerable debate about the desirability of surrogacy in any form and, more particularly commercial surrogacy.[3] This article will focus on the provisions of the Children's Act (CA) and will not explore the moralistic and other arguments associated with surrogacy. The discussion of the CA will be prefaced by a brief discussion of the nature of surrogacy and a broad overview of the legal position preceding the promulgation of the CA. This overview serves only to contextualise the discussion of the provisions of the CA and does not include a comprehensive discussion of the legal position preceding the legislative intervention. The overview will be followed by a discussion of the current legislation. Shortcomings in the current legal provisions will be explored and, finally, suggestions will be made regarding how the current legislation may be improved by the introduction of essential regulations and the design of a surrogacy specific model.

2 The Nature of Surrogacy

Surrogacy, in its most basic sense, is the situation where one woman bears a child for another.[4] The word "surrogate" means "substitute" and a woman may agree to act as a substitute mother for many reasons,

2 38 of 2005 (CA).

3 Lupton "Surrogate parenting The advantages and disadvantages" 1986 *J Juridical Sc* 148; Tager "Surrogate motherhood, legal dilemma" 1986 *SALJ* 381; Pretorius "A comparative overview and analysis of a proposed surrogate mother agreement model" 1987 *CILSA* 275; Annas "Death without dignity for commercial surrogacy" *The Hastings Centre Report* April/ May 1988 21; Lupton "The right to be born: Surrogacy and the legal control of human fertility" 1988 *De Jure* 36; Anderson "Is woman's labour a commodity" 1990 *Philosophy & Public Affairs*; Lupton "The effect of the *Baby M* case on commercial surrogacy" 1991 *TSAR* 224; Pretorius "Practical aspects of surrogate motherhood" 1991 *De Jure* 52; Satz "Markets in women's reproductive labor" 1992 *Philosophy & Public Affairs* 107; Meyerson "Surrogacy agreements" 1994 *Acta Juridica* 121; Sloth Nielsen & Van Heerden "Putting Humpty Dumpty back together again: Towards restructuring families' and children's lives in South Africa" 1998 *SALJ* 156 164-165; Louw "Surrogate motherhood" in *Commentary on the Children's Act* (eds Davel & Skelton) (2007) Ch 19; MacKenzie "Beyond genetic and gestational dualities Surrogacy agreements, legal parenthood and choice in family formation" in *Human fertilization and embryology: Reproducing regulation* (eds Horsey & Biggs) (2007) 181-204.

4 Pretorius 1991 *De Jure* 52 54.

whether financial (commercial) or compassionate (altruistic).[5] It has been suggested that the prefix "surrogate" suggests that in some way the woman who carries the child is not a "real" mother.[6] Other terms used to describe the surrogate include "hostess mother"[7] "host mother",[8] and even "plumbing".[9]

There are two main types of surrogacy agreements, total surrogacy, where the surrogate is not biologically related to the child, and partial surrogacy where her ovum is used.[10] Commissioned adoption, where an infertile couple contract with a surrogate to carry a child unrelated to either the surrogate or themselves is not permitted.[11]

Surrogacy as a means to provide families with children that are biologically related to them has received scant attention over the decades, despite it being a reality of a world in which infertility is rife and adoption processes complex.

Surrogacy became a common topic of conversation in South African households in 1987 when Pat Anthony, a Tzaneen grandmother, gave birth to her own grandchildren.[12] Anthony's daughter, Karen Ferreira Jorge then adopted the children and raised them as her own. Although surrogacy has remained a practice in South Africa, the sensationalism surrounding the birth of the Ferreira Jorge triplets has remained unparalleled since. Despite this, a wealth of articles on surrogacy followed the birth.[13]

3 The Legal Position Regarding Surrogacy Before Promulgation of the Children's Act

The Children's Status Act[14] (CSA) became operative less than two weeks after the Ferreira Jorge triplets were born.[15] This legislation provided that the gestational mother and her husband, where he consented to the artificial insemination, were the parents of a child born of artificial insemination using donor sperm or eggs.[16] By implication therefore, the gestational mother and, in the presence of spousal consent to the insemination, her husband, would be the parents of any child born of surrogacy. The CSA was not designed to deal with surrogacy and thus the unique nature of such arrangements was not considered in drafting the

5 Lupton 1986 *J Juridical Sc* 148.
6 Meyerson 1994 *Acta Juridica* 121.
7 SALC *Project 65 Report on surrogate motherhood* (1993) par 8.2.13 ff.
8 Pretorius 1991 *De Jure* 52 57.
9 Anderson 1990 *Philosophy & Public Affairs* 83.
10 See Lupton 1986 *J Juridical Sc* 148.
11 Meyerson 1994 *Acta Juridica* 121 123.
12 Pretorius 1991 *De Jure* 52 53, 55-59.
13 *Supra* n 12.
14 82 of 1987.
15 Pretorius 1991 *De Jure* 52 58.
16 S 5(1)(a) CSA.

legislation. The consequence was that the effect of the CSA was to attribute parenthood to a mother who never intended to keep the child and to a father whose involvement was minimal at best. This seems untenable.[17]

In instances where donor sperm is used for artificial insemination, the donor's rights are terminated by legislation.[18] This termination of the donor's rights might effectively mean that the husband of a surrogate mother may be unable to rebut the *pater est quem nuptiae demonstrant* presumption.[19]

The only means by which parents commissioning a baby through the surrogacy process (the commissioning parents) could acquire parental rights and responsibilities in respect of such a child was to adopt it through the normal channels.[20] This process was fraught with its own difficulties.[21]

Commissioning parents may prefer surrogacy arrangements to adoption for a number of reasons, *inter alia*, because: The nine month period of gestation associated with the pregnancy of a surrogate may be far shorter than the waiting period associated with an adoption; surrogacy allows for the possibility that one or both of the commissioning parents may be biologically related to the child; and commissioning parents are not subject to the age limits associated with adoptive parents.[22]

Thus, until the advent of the CA, surrogacy was regulated indirectly by three pieces of legislation that were designed for other purposes: The Human Tissue Act[23] (HTA) and its regulations, the Child Care Act[24] and the CSA.[25] These pieces of legislation were not ideal for a number of reasons, not least because the HTA was very restrictive in that it provided, *inter alia*, that only married women could be artificially inseminated or fertilized *in vitro*, effectively excluding unmarried women from acting as surrogates.[26]

There existed a patent need for legislative intervention. Responding to this need, the South African Law Commission (SALC),[27] as it then was, conducted research on surrogacy. This research resulted in a report and

17 Pretorius 1991 *De Jure* 52 58.
18 S 36 HTA.
19 Pretorius 1991 *De Jure* 52 58-59.
20 In terms of the Child Care Act 74 of 1983. Now in terms of ch 15 CA.
21 Lupton 1991 *TSAR* 224 230; Pretorius 1991 *De Jure* 52 59.
22 Lupton 1988 *De Jure* 36; Pretorius 1991 *De Jure* 52 54; Mills "Certainty about surrogacy" 2010 *Stell LR* 429.
23 65 of 1983.
24 74 of 1983.
25 82 of 1987. See Pretorius "Surrogate motherhood: A detailed commentary on the Draft Bill" 1996 *De Rebus* 114.
26 See case study Pretorius 1991 *De Jure* 52 61; Lupton 1988 *De Jure* 36 149.
27 Now the South African Law Reform Commission: s 5 Judicial Matters Amendment Act 55 of 2002.

a draft Bill.[28] The SALC draft Bill predated the Constitution[29] and required scrutiny regarding constitutionality.[30] Surrogacy relates to such rights as the right to procreate,[31] the right to make decisions on health and body, etcetera.[32] The Constitution does not directly protect the right to procreate although the rights to equality, privacy, religion, belief and opinion may indirectly protect this right.[33] Certainly, both the Cairo Declaration of the International Conference on Population and Development, 1994 and the African National Congress's National Health Plan for South Africa[34] support the right to freedom of procreative choice.[35]

A number of academics wrote about the draft Bill,[36] however it is sufficient here to note some of the observations made by Pretorius. In the introduction to her review of the draft Bill, Pretorius indicated that despite some public law dimension, surrogacy relates mostly to the private law sphere and stressed the importance of the advisory role of the legal professional.[37] She demanded that the legislator should provide for the status of the child and for abortion[38] and called upon the legislature to review the Canadian contract law model, which stresses careful scrutiny of commissioning parents and the need to emphasise the welfare of the child.[39]

In Pretorius' in-depth examination of the draft Bill she, *inter alia,* called upon the legislator to: Limit surrogate mothers to adults;[40] permit surrogacy in circumstance where the surrogate has given birth by means such as Caesarian section;[41] and to allow unmarried persons to act as surrogates.[42] She criticised the inclusion of a definition of marriage that would exclude lesbian and homosexual couples;[43] and the limitation of commissioning parents to married infertile couples.[44] She berated the Bill for limiting surrogacy to cases of full surrogacy, whilst recognising the

28 *Report on surrogate Motherhood Project 65* (1993).
29 Constitution of the Republic of South Africa, 1996.
30 One such examination was conducted by Pretorius 1996 *De Rebus* 114.
31 See Lupton 1991 *TSAR* 224 230 on the US procreation arguments.
32 S12 Constitution.
33 Ss 9, 14 & 15 respectively.
34 www.un.org/popin/icpd/conference/bkg/egypt/html (accessed 2012-12-6); www.imsa.org.za/PB7R%20ANC%20HEALTH%20PLAN%201994.pdf (accessed 2012-12-6).
35 Pretorius 1996 *De Rebus* 114.
36 *Inter alia,* Pretorius *ibid;* Clark "Surrogate motherhood: Comment on the South African Law Commission's report on surrogate motherhood (Project 65)" 1993 *SALJ* 769.
37 1996 *De Rebus* 115.
38 *Idem* 116.
39 *Ibid.*
40 *Ibid.*
41 *Idem* 117.
42 *Ibid.*
43 *Idem* 116.
44 *Idem* 117.

difficulties exacerbated in cases of partial surrogacy.[45] Pretorius welcomed the inclusion of clause 2 which introduced the requirement for a valid legal agreement to precede any medical procedures pursuant to surrogacy arrangements as adding certainty to the relationship and offering the various parties an opportunity to regulate important aspects of the relationship not specifically legislated for.[46] She noted that the confirmation of such agreements by the court would be premised upon a comprehensive screening of the surrogate and commissioning parents as well as the consideration of the welfare and best interests of the child.[47] Pretorius called for medical expertise to be relied upon by the court in determining the time after confirmation within which the artificial insemination should take place.[48]

In clause 8, the draft Bill stated that where a validly confirmed agreement is present, the surrogate's parental rights are terminated on birth of the child.[49] In the absence of such a validly confirmed agreement, the *status quo* is maintained.[50] Pretorius also welcomed clause 9 which provided for abortion. She opined that the final decision regarding an abortion should lie with the surrogate.[51]

Pretorius stressed that in cases of commercial surrogacy the possibility of exploitation is increased.[52] Clause 10 purported to prevented commercial surrogacy by limiting payments to the surrogate. Likewise clauses 10 to 12 limited fees payable to brokers or agencies. Pretorius called for a total ban on surrogacy agencies and on advertisement of surrogacy services.[53] Finally, Pretorius welcomed both the privacy provisions as well as the introduction of penalties for the artificial insemination of a surrogate in the absence of compliance with legislative provisions.[54]

The Bill thus envisaged that surrogacy agreements would regulate the relationship between the intended parents and the surrogate mother, and that the parties would comply with legislative imperatives. The legislative requirement for a formal agreement would prevent informal and verbal arrangements. The manner of creation of valid agreements and the legal implications of such an agreement for the parties would be legislatively regulated.

Despite the existence of the report and the draft Bill, it was not until the CA, that surrogacy was legislated for. This legislative intervention is welcome as surrogacy has been a troubled issue veiled in uncertainty for

45 *Idem* 118-119.
46 *Idem* 119.
47 *Ibid*.
48 *Ibid*.
49 Clause 8(1).
50 Clause 8(2).
51 Pretorius 1996 *De Rebus* 114 121.
52 *Ibid*.
53 *Ibid*.
54 *Ibid*. Clauses 11 & 12 Draft Bill respectively.

a very long time. As will appear from the discussion that follows, chapter 19 of the CA is not however, the cure-all that was hoped for.

4 Surrogate Motherhood Agreements in South Africa and Chapter 19 of the CA[55]

4 1 The Surrogate Motherhood Agreement

All surrogacy arrangements in South Africa must now be the subject of a valid, written surrogate motherhood agreement, the provisions of which, together with the CA, regulate the surrogacy arrangement.[56] The surrogate motherhood agreement has been described as "a contract of a special kind"[57] and is defined as:

> [a]n agreement between a surrogate mother and a commissioning parent in which it is agreed that the surrogate mother will be artificially fertilised for the purpose of bearing a child for the commissioning parent and in which the surrogate mother undertakes to hand over such a child to the commissioning parent upon its birth, or within a reasonable time thereafter, with the intention that the child concerned becomes the legitimate child of the commissioning parent.[58]

The effect of the common law maxims *mater semper certa est* and *pater est quem nuptiae demonstrant* can thus be altered by a validly concluded and confirmed surrogate motherhood agreement.[59] Consequently a surrogate motherhood agreement actively aims to thwart the effect of certain common law and legislative rules concerning families and children.[60] Chapter 19 of the CA aims to systematically order these surrogate motherhood agreements, but, being new law in South Africa, the provisions are surrounded by uncertainty.[61] Given the far-reaching consequences of surrogacy agreements for all parties involved, it is imperative that the utmost care is taken in drafting the agreement and compiling supporting documentation.[62]

---

55 In approaching the CA, a predominantly contract law model has been assumed. A detailed discussion follows infra at 6 2. For the purpose of the remainder of this article only situations of formal surrogacy, as regulated by the CA, will be discussed. For a discussion of informal surrogacy see Sloth Nielsen & Van Heerden 1998 *SALJ* 156 164; Louw 19-3.

56 S 292 CA.

57 *Ex parte WH* 2011 6 SA 514 (GNP) 530D.

58 S 1 CA (*sv* "surrogate motherhood agreement").

59 S 297 CA.

60 Annas *Hastings Center Report* April/May 1988 21; Lupton 1991 *TSAR* 224 225.

61 *Ex parte WH* 2011 6 SA 514 (GNP) 524B-C.

62 See *In re confirmation of three surrogate motherhood agreements* 2011 6 SA 22 (GSJ) (hereinafter referred to as "the *Confirmation* case") 25H-I regarding consequences for legal practitioners who do not draft these applications with the due care and respect for judicial procedure.

The legislative regulation of the content, conclusion and confirmation of surrogate motherhood agreements is essential for various reasons, *inter alia*, to give effect to the best interests of the child(ren)[63] born of surrogacy;[64] to minimise the risks attached to surrogacy arrangements; to give effect to the wishes of all the parties involved; and to clarify the parental responsibilities of the parties to the agreement.[65]

The surrogate motherhood agreement is regarded as being so complex that the general principles pertaining to the law of contract alone were deemed inadequate to regulate it.[66] For this reason, chapter 19 of CA was needed to stringently regulate the parental responsibilities that flow from the validly concluded agreement. Chapter 19 also attempts to provide guidelines regarding the drafting of the terms of such an agreement. Sadly however, CA fails to establish guidelines regarding supporting documentation that should be submitted to the court in support of the confirmation of such agreements. The South African courts recently attempted to clarify matters in *Ex parte: WH*[67] and *In re confirmation of three surrogate motherhood agreements*[68] by aspiring to develop a standard practice for confirmation applications.[69] Pursuant to its decision in the latter case, the South Gauteng High Court published a practice directive relating to the confirmation of surrogate motherhood agreements.[70]

63 Where reference is made to "the child" it must be read to include the plural as well.

64 S 28(2) Constitution enshrines such protection, Louw 19-7.

65 Louw 19-7.

66 SALC *Project 65* par 8.3.

67 2011 6 SA 514 (GNP). In this case the court confirmed that homosexuality should not exclude a couple from qualifying as commissioning parents, as this would be unfairly discriminatory and thus unconstitutional (526F-527A).

68 2011 6 SA 22 (GSJ). The court had to deal with an attorney who "copied and pasted" three surrogate motherhood agreements and brought them before the urgent court. The agreements were not confirmed because the court did not deem them urgent and the grossly negligent manner in which the applications were drafted and compiled most likely hindered the case even further. The matter was postponed for three months during which time the attorney opened new case files and brought the matter before three more judges in an attempt to have the agreements confirmed. In their judgment Wepener J & Victor J referred the matter of punishing the attorney to the Law Society of the Northern Provinces (24I-25F).

69 Nöthling Slabbert *SA J Bioethics and Law* June 2012 28.

70 Practice Directive 05 of 2011 of the South Gauteng High Court, *Re: Application for Confirmation of Agreements in terms of Section 295 Children's Act* 2011-02-16.

4 2 Legislative Requirements for a Validly Concluded and Confirmed Surrogate Motherhood Agreement[71]

Section 292 of the CA provides the formal requirements for a valid and enforceable surrogate motherhood agreement.[72] This section provides that the agreement between the parties must be in writing and signed by all involved;[73] it must be concluded in the Republic of South Africa;[74] at least one of the commissioning parents (or the commissioning parent if a single person)[75] and the surrogate mother (and her "husband"[76] or partner, if any) must be domiciled in the Republic at the time of concluding the agreement;[77] and, most importantly, the surrogate motherhood agreement must be confirmed by the high court within the area of jurisdiction in which the commissioning parent is domiciled or habitually resident.[78] A lack of compliance with these fundamental requirements will invalidate the agreement and render it unenforceable between the parties.[79]

The following substantive requirements are entrenched in the remainder of Chapter 19:

(a) The agreement must contain the written consent of the partner or spouse of the commissioning parent, where the commissioning parent is in a permanent relationship or married.[80] This requirement applies *mutatis mutandis* to the partner or spouse of the surrogate mother.[81]

(b) The surrogate motherhood agreement must include "adequate provisions for the contact, care, upbringing and general welfare of the child that is to be born in a stable home environment" and should further include detailed provisions concerning the child's position in the

71 It must be borne in mind throughout, that the best interest of the child and the individual circumstances of the parties must always be considered by the courts when adjudicating these matters (s 295(e) CA).

72 In *Ex parte WH Supra* (523B-C, 530G-H) the court distinguishes between the "formal" (s 292 CA) and the "substantive" legislative requirements (found in the remainder of ch 19 CA).

73 S 292(1)(a) CA.

74 S 292(1)(b) CA.

75 S 292(1)(c) CA. Where reference is made to "the commissioning parents" it must be read to include the singular as well.

76 Louw's argument (19-10) that the term "husband" is unconstitutional and discriminatory is supported and it is suggested that the CA be amended to read "spouse".

77 S 292(1)(d) CA. In terms of s 292(2) CA this requirement may however be set aside by a court if "good cause" can be shown.

78 S 292(1)(e) CA.

79 S 292(1) CA. For the effect of an invalid surrogate motherhood agreement see the discussion infra at 5 3.

80 S 293(1) CA.

81 S 293(2) CA. If such consent is unreasonably refused (and the spouse is not the genetic parent of the child), the Court may still confirm the agreement in terms of s 293(3) CA. See further Louw 19-10 – 19-11. This consent requirement is even more crucial where the partner and the surrogate mother are in a "permanent life-partnership" and he is also the biological parent of the child (see further Louw 19-9).

eventuality of the death, divorce or separation of the commissioning parents prior to the birth of the child.[82]

(c) The commissioning parent, or in the case of two commissioning parents, one or both must be a biological parent of the child.[83]

(d) The sterility of the commissioning parent(s) must be irreversible.[84]

(e) The artificial fertilisation of the surrogate mother is obligatory, but may only be performed in the eighteen-month period after a high court has confirmed the surrogate motherhood agreement.[85]

(f) Importantly, the agreement must stipulate that no compensation or promise to compensate (in cash or otherwise) has been made by any party to the surrogate mother.[86] This excludes compensation related to the fertilisation, pregnancy and birth or the confirmation of the surrogate motherhood agreement;[87] loss of earnings of the surrogate mother as a result of the conclusion of the agreement;[88] or insurance to cover death or disability suffered by the surrogate mother as a result of the pregnancy.[89]

Certain requirements and formalities concern the parties themselves:

(a) The commissioning parents should be competent, as required by the CA, to enter into the agreement;[90] be appropriate individuals to act as parents to the child;[91] and appreciate and recognise the legal consequences of the conclusion of the agreement and the rights and responsibilities emanating from it.[92]

(b) The surrogate mother should be competent, as required by the CA, to enter into the agreement;[93] constitute a "suitable" person to act as a surrogate;[94] appreciate and recognise the legal consequences of the agreement and her rights and responsibilities in terms thereof;[95] not use surrogacy to acquire compensation but should be motivated by a selfless drive to help others;[96] and finally, have a documented history of one or more pregnancies, "viable" deliveries and living children of her own.[97]

---

82 S 295(d) CA.
83 S 294 CA. It has been argued that s 294 CA is unconstitutional as it infringes on the infertile person's rights to freedom, dignity and decisions regarding reproduction. See Louw 19-13 in this regard.
84 S 295(a) CA.
85 S 296(a), (b) CA.
86 S 301(1) CA.
87 S 301(2)(a) CA.
88 S 301(2)(b) CA.
89 S 301(2)(c) CA.
90 S 295(b)(i) CA. Louw 19-15 presumes that this refers to the *domicilium,* consent, genetic link and sterility requirements as set out in ch 19 CA.
91 S 295(b)(ii) CA.
92 S 295(b)(iii) CA.
93 S 295(c)(i) CA. Louw (19-16) opines that this section refers to the fact that the surrogate mother should be an adult of sound mind and the *domicillium* and consent requirements of her and her spouse.
94 S 295(c)(ii) CA.
95 S 295(c)(iii) CA.
96 S 295(c)(iv), (v) CA.
97 S 295(c)(vi), (vii) CA.

Guidelines for the termination of the pregnancy or of the surrogate motherhood agreement and the effect of such termination are provided by the CA.[98]

5 Problematic Requirements in terms of Chapter 19 of the CA

The two most recent cases relating to surrogate motherhood agreements[99] demonstrate that the CA is unclear regarding what is required of the parties to such agreements. Furthermore, our courts have not yet ruled on chapter 19 extensively. Some of the requirements that give rise to uncertainty will now be considered in more detail.[100] The discussion will incorporate a brief evaluation of recent case authority and recommendations regarding the promulgation of regulations to the CA designed to address persistent uncertainties.

In relation to the general legislative requirements to be met by parties applying for confirmation of a surrogate motherhood agreement, the North Gauteng High Court has ruled that the affidavit presented to the court in support of the application should contain "all factors set out in the Act together with documentary proof where applicable".[101] An example would be documentary proof regarding the sterility of the commissioning parent(s) in terms of section 295(a). Such proof may take the form of a letter from a specialist medical professional stating the exact cause of the irreversibility of the sterility of the commissioning parent(s). The court emphatically stated that for an application for confirmation of the agreement to succeed, every aspect regarding the application and the agreement required by the CA must be proven to the satisfaction of the court. The question that remains is what proof of satisfaction of legislative prerequisites will be required.

The court has warned that its role is not simply to "rubber stamp" agreements.[102] The court views its role as upper guardian of all children as overriding and, as such, expects to be fully apprised of certain facts pertaining to the parties and their circumstances.[103] These facts and circumstances can only be made known to the court through the presentation of evidence, the veracity of which can then be evaluated by the court.

98 Ss 298-300 CA. See the discussion on breach of contract infra at 5 7.
99 *Ex parte WH* and the *Confirmation* case *Supra*.
100 Please note that this discussion will deal only with certain aspects of the legislation that appear potentially problematic and does not purport to be a complete discussion of the limitations in the legislation.
101 *Ex parte WH Supra* 531D.
102 *The Confirmation case Supra* 30F. See further 25E-F for a warning against attempting to avoid proper procedure.
103 *The Confirmation case Supra* 28B-29D.

The current lack of regulations pertaining to chapter 19[104] hinders the process of compiling an application for the confirmation of a surrogate motherhood agreement, as it remains unclear what constitutes sufficient evidence of compliance with the requirements.

Some of the requirements set out in the CA are themselves problematic. A brief exposition of these follows.

5 1 Requirements Relating to the Parties

The first problematic requirement relates to the individuals who are required to be parties to the agreement. Clearly, the commissioning parent[105] and the surrogate mother[106] must be parties, but the Act also requires the written consent of the commissioning parent and surrogate mother's respective partners if a "permanent relationship" is present on the facts. Where the commissioning parent is married or in a permanent relationship, it is unclear whether both should be identified as commissioning parents or whether the consent requirement is automatically met when they both become parties to the agreement.[107] The concept "permanent relationship" is vague and it has been opined that the courts will interpret the concept widely.[108] This matter has not yet been judicially considered and, for the sake of clarity, a regulation should stipulate what exactly parties should prove in this regard. Guidelines to determine the permanence of a relationship; who should be identified as a commissioning parent; and exactly who should consent should be provided for in a regulation to chapter 19.

The potential for interesting scenarios to develop through medical and technological advances should be provided for, given that the law responds slowly to social change. Presently, the technology already exists to create a human embryo from the sperm of a man and the ova of two women.[109] Implementing such technologies in relation to surrogacy might still be some way off, but it begs the question whether the CA is equipped to deal with such complex situations.

---

104 When presenting a version of this article at the Society of Law Teachers of Southern African Conference in July 2012, Sloth Nielson provided crucial information regarding the promulgation of these suggested regulations. She was tasked with the drafting of regulations to the CA, but a governmental "turf war" obstructed her efforts relating to regulations to ch 19 CA. Reportedly, the Departments of Social Welfare and Development, Health and Justice and Constitutional Development are all uncertain as to within whose jurisdiction matters regarding surrogacy fall. It is obvious that before such uncertainty is remedied none of the proposed, and much needed, regulations will be promulgated.

105 S 293(1) CA.

106 S 293(2) CA.

107 Louw 19-9 – 9-10.

108 *Idem* 19-9.

109 http://m.news24.com/news24/SciTech/News/New-embryo-methods-should-be-allowed-20120612 (accessed 2012-06-13).

5 2 Requirements Relating to Care

The surrogate motherhood agreement must include sufficient provisions regarding the care and stable home of the child.[110] The South African courts have not yet stipulated what should be incorporated in "care" clauses in order to be deemed sufficient. Furthermore, the CA does not explain what constitutes a stable home, or by whom the existence thereof should be determined. Clearly this section would benefit from the promulgation of a regulation providing for an appropriate screening process.[111] Louw addresses the poor drafting of section 295(d), stating that by "anticipating the eventualities" related to this section an attempt can be made to ensure the effectiveness and enforceability of the surrogate motherhood agreement as well as facilitate its confirmation.[112] The suitability of the commissioning parents is intrinsically linked to this care-requirement.

In addition, section 295(d) requires that the agreement should include provisions regarding the child's position in the event of the death of the commissioning parents or their divorce or separation before the birth of the child. Here it is important to note that in the *Confirmation* case the court specifically required a psycho-social analysis of the suitability of *the person designated by the commissioning parent* in the event of his or her death.[113] This requirement clearly illustrates how seriously the court considers an application of this nature.

5 3 Requirements Relating to Fertilisation

The requirement relating to the artificial fertilisation of the surrogate mother creates various problems and uncertainties. These will be discussed briefly.

5 3 1 Time Frame

Firstly, the provision that the fertilisation should take place within an eighteen month period is troublesome as the process of artificial fertilisation can be time consuming and is often not successful on a first or even second attempt.[114] The relatively brief timeframe within which the fertilisation must be effected may necessitate the drafting and confirmation of a further surrogate motherhood agreement before the

110 S 295(d) CA. Louw (19-18) correctly questions why the legislature deemed "care" insufficient.

111 Louw 19-18.

112 *Ibid*.

113 30B, our emphasis.

114 Pretorius 1996 *De Rebus* 114 115. Louw (19-20) opines that this eighteen month restriction was likely put in place to protect the parties from radically altered circumstances. The time restriction is necessary, but extending it to 24 months could lead to a more equitable situation in practice. Cl 7 Draft Bill on Surrogate Motherhood prescribed a twelve month period, so some leeway has been granted in the CA, but it is argued that eighteen months is still too short.

process is completed. This will have severe cost implications for the commissioning parents.[115]

5 3 2 Confirmation of the Surrogate Motherhood Agreement

More problematic and unsettling than the cost implications associated with this requirement is the fact that it appears that in practice, parties are ignoring the requirement that court confirmation of the surrogate motherhood agreement must precede the artificial fertilisation of the surrogate mother.[116] The law is thus being flouted, either out of ignorance or disrespect. This may be to the detriment of the parties or the child.

Surrogate motherhood agreements that are not confirmed by a court are invalid and unenforceable.[117] If there is no valid agreement, either because no agreement was entered into or the agreement is unconfirmed, the child born of surrogacy will be considered the child of the surrogate mother and her partner, if any.[118] The commissioning parents would then have to adopt the child.[119] This process could become complicated when one (or both) adoptive parent(s) is biologically linked to the child, as problems are associated with adopting one's own child.[120] The fact that some commissioning parents ignore this crucial requirement for prior confirmation of the agreement, because they will do anything to have a child as soon as possible, is short-sighted. When the parentage of a child comes into question it could have disastrous effects on the child in question.

The lack of a confirmed surrogate motherhood agreement before fertilisation raises another troubling issue, this time in relation to the conduct of medical professionals. Section 303(1) emphatically prohibits any person from assisting in or effecting the artificial fertilisation without a confirmed surrogate motherhood agreement.[121] Failure to comply is a criminal offence[122] punishable by a fine, a maximum of twenty years imprisonment or both.[123] Without clarity on the degree of fault

115 Some report amounts as high as R450,000 for the entire process. See *Daily News* (2011-01-17) 11.
116 Informal interview with Dr Magriet Coetsee-Spies, 2012-08-24.
117 S 292(1)(e) CA.
118 S 297(2) CA. See further Louw 19-19 n 6. Meyerson 1994 *Acta Juridica* 121 137 argues that invalidating the contract provides no protection to the surrogate mother.
119 S 294 CA.
120 In terms of s 17 Child Care Act this was prohibited (Pretorius 1991 *De Jure* 52 59). See s 231(1)(c), (d), (7)(a) CA in this regard. A problem still arises where the commissioning mother is the biological parent and she does not have guardianship of the child. Although the CA improves the situation where the father is a biological parent, uncertainty still reigns where the mother or both parents are genetically linked to the child.
121 See Louw (19-19 - 19-20) for a discussion pertaining to the required culpability of medical professionals.
122 S 305(1)(b) CA.
123 S 305(1)(b) CA read with s 305(6), (7) CA.

necessary for criminal conviction, medical practitioners should be made aware of this provision and the repercussions of their failure to comply.[124] Before promulgation of the CA, doctors were often unwilling to assist couples due to their fear of attracting liability.[125] Regretfully, since surrogacy has been legalised and regulated these medical practitioners seem to be blasé about the possibility of criminal prosecution.[126] The effect of this disregard, for what can be considered a core legislative requirement relating to the surrogacy process, is to undermine the integrity of legal, medical and social service professionals. "Pressure from clients"[127] should not under any circumstances justify the shirking of legal, ethical or professional responsibilities.

5 3 3 Origin of Gametes

The court specifically required full information regarding the source of the gametes, without necessarily identifying the donors.[128] This would be necessary to establish whether or not the surrogate mother is genetically related to the child, which will affect the rules regulating the termination of the agreement.

The topic of egg donation is closely related to surrogacy as the two processes often go hand-in-hand. Chapter 8 of the National Health Act[129] (NHA) and its regulations[130] relate to the control of the use of human gametes. These provisions did not specifically regulate surrogacy but also do not specifically excluded surrogacy from their ambit.[131] Section 60(4)(a) of the NHA declares it an offence to receive any form of reward, financial or otherwise, for the donation of a gamete by the person donating it, "except for the reimbursement of reasonable costs incurred by him or her to provide such donation". Regulation 4 of the Regulations Relating to Artificial Fertilisation of Persons[132] states that the donor may only be reimbursed "for any reasonable expenses incurred by him or her in order to donate a gamete" as contemplated in section 60(4)(a) of the NHA.

What can be deemed as "reasonable" in this context remains ambiguous and the regulation fails to expand upon or clarify the provision in any meaningful way. The *2008 Guidelines for Gamete*

124 On whom this "duty" rests to inform medical professionals of their responsibilities in terms of ch 19 CA is another unanswerable question. This issue is however directly linked to the governmental jurisdiction debate (*Supra* n 104).
125 Pretorius 1996 *De Rebus* 114 115.
126 Interview with Dr Coetsee-Spies *Supra*.
127 See in this regard the *Confirmation* case 25F par 12.
128 *Ex parte WH Supra* 329F.
129 61 of 2003.
130 Regulations published under GN R175 GG 35099 2012-03-02.
131 Nöthling Slabbert *SAJBL* June 2012 29. She views the need for regulations to ch 19 CA from the vantage point that the NHA and its regulations only regulate artificial insemination, but not as it relates to surrogacy.
132 *Supra* n 130.

Donation of the Southern African Society of Reproductive Medicine and Endoscopic Surgery[133] deals directly with this matter, stating that any compensation should reflect the time, inconvenience, expenses (travel, loss of income and childcare costs), physical and emotional demands and risks of the donor associated with the egg donation. It is also stated that the tariff compensated should minimise "the possibility of undue inducement of donors" or any implication that the compensation is for the gametes themselves.[134] The proposed compensation is R5,000.00 to R6,000.00 and the payment of compensation exceeding R10,000.00 "should only be paid in exceptional circumstances."[135]

Various red flags are immediately raised: what constitutes "exceptional circumstances"; would paying compensation in excess of R6,000.00 be acceptable even if not advised; and most importantly, why was a matter as crucial as this not sufficiently addressed in the regulations to Chapter 8 of the NHA? In contrast, the South African Medical Research Council has stated that where eggs are donated for the purposes of medical research no compensation may be paid for such donation.[136] The problem with these guidelines is that they are just that, unenforceable guidelines. It is thus advisable that the legislature should rethink and clarify its position relating to compensation for the donation of eggs used for human reproduction.

5 3 4 Intermediaries

Couples seeking to use surrogacy as an option to conceive a child appear to first approach egg donation clinics, surrogacy agents or fertility specialists for assistance.[137] They are then put in contact with social workers and psychologists who perform suitability assessments on the parties involved.[138] These prospective commissioning parents often first hear about the legal requirements through these agencies whose role must thus be carefully scrutinised and more closely monitored. Some agencies suggest the possibility of using both donor egg and sperm, which is not permitted by the CA and is tantamount to commissioned adoption.[139] Providing the egg donor with a gift is also encouraged: "This is please not to be [sic] an excessive gift, but just a small gesture to express gratitude for this amazing deed".[140] It is submitted that such a

133 http://www.fertilitysa.org.za/EggDonation/2008%20GUIDELINES%20FOR%20GAMETE%20DONATION.doc (accessed 2012-09-13).
134 *Ibid.*
135 *Ibid.*
136 *Guidelines on Ethics for Medical Research: Reproductive Biology and Genetic Research* Book 2 (2002) par 2.8 (http://www.sahealthinfo.org/ethics/ethicsbook2.pdf (accessed 2012-09-19).
137 Interview with Dr Coetsee-Spies *Supra*.
138 *Ibid.*
139 http://baby2mom.co.za/page/11588/IntendingParentsViaEggDonation#%2FPage%2F11589%2F What-Does-Egg-Donation-Cost%3F (accessed 2012-08-17). See also Nöthling Slabbert *SAJBL* June 2012 31.
140 http://baby2mom.boltcms.com/Page/11588/Recipients#%2FNews%2F188%2FEgg-Donor-Gift (accessed 2012-08-17).

"gift" equates compensation and may been viewed as an incentive, seriously bringing the ethics of these agencies into question. The practice of giving a gift could thus be construed as an attempt to sidestep the compensation guidelines.

The CA does not outlaw the practice of informal surrogacy between individuals such as friends or family members, however, where parties are contemplating informal surrogacy and intend to approach a family member or friend to act as surrogate, a surrogate motherhood agreement is still advised. Formalising the process will clarify the rights and responsibilities of all involved and thus aid in protecting the rights of each party.

Finding a surrogate is a contentious issue. Agencies inform commissioning parents that it is desirable that they find their own surrogate mother but indicate that when they, the agency, put the commissioning parents in contact with a potential surrogate there is no waiting period.[141] These statements are irreconcilable. Agencies have lists of women readily and immediately available to act as surrogate mothers to total strangers. The motives of both the potential surrogate and the agency "brokering" the agreement must be questioned, as very few women would volunteer to act as a surrogate to strangers if no fee was offered for their services and discomfort.[142]

5 4 Compensation

The court will only confirm a surrogate motherhood agreement if it is certain that no compensation or promise to compensate, in cash or otherwise, has been made by any party involved to the surrogate mother.[143] The courts have thus greatly expanded upon the information to be provided in order to ensure that the proposed surrogacy is not of a commercial nature.

Full particulars should be set out in the founding affidavit of how the commissioning parents became aware of the surrogate mother and exactly why she is willing act in this capacity.[144] The affidavit should further include information regarding any and all agreements between the surrogate mother and any intermediary, such as an agency or clinic, and between such intermediary and the applicants.[145] Details and documentary proof of the payment of any compensation, either to the surrogate mother, an intermediary, donor, clinic or any third party involved should be provided to the court, regardless of the service being

141 http://baby2mom.boltcms.com/Page/11617/SurrogacyProgram#%2FPage%2F11838%2FFAQ-for-Intending-Parents-Via-Surrogacy- (accessed 2012-08-17).
142 Meyerson 1994 *Acta Juridica* 121 138.
143 S 301(2) CA.
144 *Ex parte WH Supra* 529C-D.
145 *Idem* 531G.

compensated for or by whom such compensation is to be paid.[146] The affidavit should reflect that no facilitation fee was paid to any person for the introduction of the surrogate mother to the commissioning parent.[147] A detailed and specific "list of surrogacy expenses with sufficient specificity"[148] should also be provided. This is important so as to prevent the payment of compensation disguised as a generic payment for expenditure.[149]

The court required these details of the agency after it became aware that the agency was facilitating introductions of surrogates and commissioning parents in order to ensure that no middleman was compensating the surrogate mother.[150] The court's requirement of this information stems from the role it serves to protect women from the exploitation that could flow from such intermediaries' involvement.[151] Stricter regulation of the role, if any, that surrogacy agencies are to play is of the utmost importance if the legislative requirement of purely altruistic surrogacy is to be respected. The prevailing socio-economic landscape of South Africa and the hardships faced by certain groups of women could potentially create a situation in which agencies could exploit women.[152]

Commercial surrogacy has been equated to slavery, but in contrast it has been argued that paying a fee for a service rendered "is merely consistent with liberal capitalist economies".[153] Pretorius[154] contends that "most South African authors" writing on surrogacy between 1982 and 1991 favoured the banning of commercial surrogacy and most later authors agree or fail to comment on the matter.[155] The view that agencies are purely motivated by the profit-potential is widely held.[156] In the USA surrogacy brokers exploit infertile couples and surrogates to make large profits. It is often found that these agencies are not run by medical practitioners, but by lay people, who are purely motivated by greed.[157] In South Africa the possibility for exploitation is great and some believe that, in the best interests of all involved, these agencies and advertisements for them should be banned outright.[158] Pretorius equates keeping lists of potential surrogates to commercial surrogacy[159] and we support this view.

146 *Idem* 531F.
147 *Idem* 528H-I.
148 *Idem* 521D.
149 *Idem* 521C-D.
150 *Idem* 519C.
151 Lupton 1991 *TSAR* 224 225, 228.
152 *Ex parte WH Supra* 528F-G.
153 Clark 1993 *SALJ* 774.
154 1991 *De Jure* 56.
155 Meyerson 1994 *Acta Juridica* 121 being the exception.
156 Annas *Hastings Center Report* April/May 1988 22, 23; Lutpon 1991 *TSAR* 227.
157 Pretorius 1996 *De Rebus* 114 121.
158 *Ibid.*
159 *Ibid.*

The fact that newspapers refer to "wombs for hire",[160] "babymakers" and "Rent-a-womb"[161] is not helping to remedy the erroneous perspective that it is possible to simply pay a woman to carry a child and that parenthood will automatically follow for the commissioning parent.[162] It can also be argued that such reports create the impression amongst potential surrogate mothers that they can earn large sums of money by carrying children for others.

Section 303(2) of the CA specifically prohibits the advertising of women's willingness to act as surrogate mothers if done for profit or with any view to compensation. The implication thereof is that it is illegal to arrange surrogate motherhood agreements or relationships on a commercial basis.[163] The question can then be posed as to whether or not, what most South African surrogacy agencies do in the course and scope of their everyday operations falls within the strict purview of the law. They receive fees from commissioning parents that are directly related to the surrogacy practice and are thus making money from their facilitation of surrogate motherhood agreements.

In light of section 12(2)(a) of the Constitution protecting a woman's right to make decisions regarding reproduction, it could possibly be argued that the ban on commercial surrogacy is unconstitutional, but it is contended that such an infringement does in fact reflect current public policy. Commercial surrogacy has rightfully been condemned to a death without dignity and attempts to revive it or practice it illegally "would not be in the best interests of children, families or society".[164] Therefore, regulations to Chapter 19 of the CA should expressly stipulate in detail what surrogacy and egg donation agencies may and may not do. The safe-guarding of altruistic surrogacy as the sole form of surrogacy practiced, is not ensured by the protection currently provided under Chapter 19 of the CA.

5 5 Suitability

Possibly the most important, yet uncertain requirements are those regarding the suitability of the commissioning parents and surrogate mother to act in these respective capacities. The courts have stated that in order for them to make assessments regarding whether the parties are fit and proper persons, the applicants "must supply proper and full details regarding themselves"[165] and that nothing but the "utmost good faith" would be tolerated.[166]

160 *Sunday Tribune* (2012-01-15) 8.
161 *Sunday Times* (2009-10-25) 10.
162 Meyerson 1994 *Acta Juridica* 121 argues that the use of such rhetorical phrases deteriorates the quality of the entire surrogacy debate.
163 Nöthling Slabbert *SAJBL* June 2012 30.
164 Annas *Hastings Center Report* April/May 1988 24.
165 *The Confirmation* case *Supra* 30E.
166 *Ex parte WH Supra* 530H.

The details necessitated by the court include: the identities and full backgrounds of the commissioning parents; proof of their financial position; an exposition of their residential situation; and their criminal records[167] if any.[168] A clinical psychologist and social worker respectively, should draft expert assessment reports on the suitability and stability of the commissioning parents.[169] Here sufficient supporting documentation is of the utmost importance:

> Ultimately the court must be satisfied that the conclusions arrived at are supported by the facts. Accordingly, vague and generic allegations in this regard that fall short of supporting a conclusion may well render an application defective.[170]

A word of caution has been aimed at judges, warning them against allowing their private prejudices and preferences to influence their decisions on the suitability of parties.[171] Guidelines can be found in the prescription of an objective test when assessing commissioning parents: the court should evaluate the prospective parents' ability to provide a safe environment for the healthy and optimal development of the child, as well as their ability to provide emotional and financial care.[172] It is necessary for the court to have a wide array of information at its disposal in assessing the suitability of the commissioning parents, as their parenting of the intended child will go beyond their contractual statement of intentions.[173] Some aspects evaluated during the psycho-social assessment of commissioning parents may include the presence of existing psychological conditions, the reason for their desire to have a child, the length and stability of their relationship and whether or not the parent not genetically linked to the child will display a jealous tendency which might cause conflict in the relationship or adversely affect the child.[174] The inevitable result of surrogacy and adoption is that prospective parents are held to a higher standard of care than natural parents,[175] but this is unfortunately an unavoidable by-product of the quest to protect the child.

The court desires even more information on the surrogate mother before she will be considered fit to act in this capacity. A thorough medical report concerning the surrogate mother's health and physical suitability to bear a child is important for obvious reasons, but the report

---

167 The *Ex parte WH* case (*Supra* 531H) specifically requires an affidavit stating whether any of the commissioning parents have been charged with or convicted of a violent crime or a crime of a sexual nature.
168 The *Confirmation* case *Supra* 29B-C.
169 *Ibid.* The court refers to the "good practice" relating to adoption and that it can be fruitfully applied to surrogacy.
170 *Ex parte WH Supra* 541A-B. See further Sloth Nielsen & Van Heerden 1998 *SALJ* 156 164.
171 *Ex parte WH Supra* 529G-J.
172 *Idem* 530B-C.
173 Annas *Hastings Center Report* April/May 1988 22.
174 Interview with Dr Coetsee-Spies *Supra*.
175 Meyerson 1994 *Acta Juridica* 121 128.

should also include information on her HIV status and any other illness that could possibly be transmitted to the child. The affidavit should contain detailed information pertaining to the surrogate mother's identity, background and financial position. The court should also be furnished with an extensive report by a psychologist, and preferably also a social worker, detailing her background and psychological profile, attesting to her suitability to act as surrogate mother and explaining the psychological effect the surrogacy and giving up the newborn could possibly have.[176] Whether the surrogate mother will act in a responsible manner to ensure the healthy development of the child is yet another important aspect that should be considered during the assessment.[177]

In *Project 65* the SALC[178] recommended a strict six month screening process and counselling[179] for all parties before and after the conclusion and implementation of the agreement.[180] This provision aimed to ensure the suitable social and psychological backgrounds of all parties involved.[181] Twelve months of HIV testing for all parties was also considered to be essential before the fertilisation.[182] The screening was to be conducted by a state-run or private body "approved by legislation".[183] In light of the court's requirement of the "utmost good faith",[184] it can be argued that these assessments should preferably be done by accredited state-run entities.[185] Some surrogacy agencies "strongly recommend" specific individuals for these assessments.[186] It is questionable whether these individuals are recommended for their knowledge and understanding of the complex nature of surrogacy or because they are compliant in the issuing of positive reports against payment of their fee.

The CA, as it relates to the suitability of the surrogate mother and commissioning parent(s), is thus severely lacking as none of these recommendations, which seem crucial to the effective and successful implementation of the surrogate motherhood agreement, have been regulated. Unfortunately, the guidelines provided by the courts also fail

176 *Ex parte WH Supra* 529D-F.
177 Meyerson 1994 *Acta Juridica* 121 143.
178 As it was known then, see n 27.
179 In California some attorneys suggest group counselling throughout the period of the pregnancy, attributing a higher success rate to such an approach (Pretorius 1991 *De Jure* 52 54); Clark 1993 *SALJ* 777.
180 SALC *Project 65 Supra* par 8.2.3.
181 Louw 19-15 n 8.
182 Louw 19-15.
183 *Ad Hoc* Committee on *Surrogate Motherhood* par F6(2)-(3). For a detailed discussion regarding the screening of parties see Louw 19-15.
184 *Ex parte WH Supra* 531H.
185 This should preferably be done by a government department, body or committee in terms of guidelines set out in legislation or regulations to the CA. Louw (19-16 n 1); Meyerson 1994 *Acta Juridica* 121 143 reiterates the importance of the objectivity of such an assessment.
186 http://baby2mom.boltcms.com/Page/11617/SurrogacyProgram#%2FPage%2F11621%2FPsychologists-for-Surrogacy-Assessments (accessed 2012-08-17).

to clarify the matter. The promulgation of regulations to the CA remains an avenue to address these omissions,[187] and when, or if,[188] these are drafted care should be taken to provide details pertaining to screening panels, timeframes and aspects to be monitored. The South African Council for Social Service Professions could draft guidelines for the evaluation of parties to surrogate motherhood agreements to strive to achieve a uniform system until regulations have been promulgated.[189]

5 6 Incidentalia

It should be noted that the parties may voluntarily include certain types of clauses in their surrogate motherhood agreement.[190] The purpose thereof is purely to regulate matters not addressed by the CA and not to attempt to alter the effect of the legislative provisions. The list of possible clauses is endless, but it should be borne in mind that these *incidentalia* should be both lawful and in accordance with the prevailing *boni mores*.[191]

Some of these *incidentalia* may include clauses requiring: Social disease testing on all parties;[192] assurance provided by the commissioning parents that they will accept parentage of the child regardless of physical or mental defects or disabilities;[193] clarification of the situation where such disabilities are due to actions of the surrogate mother;[194] clarification regarding a duty of care where the surrogate mother is advised by a medical professional to terminate the pregnancy, and she chooses not to do so; policies insuring the health and life of the surrogate mother to be maintained by the commissioning parents for the duration of the pregnancy;[195] the clarification of the financial responsibilities of the parties, including provisos requiring funds to cover all anticipated expenses be kept in trust;[196] suggested courses of action in the eventuality that any of the parties wishes to terminate the pregnancy or the agreement;[197] the possibility of visitation rights, if any, of the surrogate mother;[198] confidentiality or privacy statements;[199] specifications regarding the number of embryos to be transferred at one

187 Louw 19-16.
188 See n 104 *Supra*.
189 A request for copies of any existing guidelines from this body was met with a response that they were unable to assist. Whether this was because such guidelines do not exist or because they were unwilling to make them available was unclear.
190 SALC *Project 65 Supra* par 8.3.1; Louw 19-18.
191 Louw 19-7.
192 *Ad Hoc* Committee's *Report* par F7(4)(j).
193 Pretorius 1996 *De Rebus* 114 117, 121.
194 See Louw's discussion (19-23) of such a situation.
195 Pretorius 1996 *De Rebus* 114 117; Louw 19-18.
196 Louw 19-18.
197 *Ibid*.
198 *Idem* 19-19.
199 *Ibid*. See also s 302 CA in this regard.

time;[200] and the number of attempts to transfer embryos that the parties may make throughout the course of the validity of the agreement. The CA requires provision for the care of the child if the commissioning parents pass away, separate or divorce before the birth of the child.[201] However a similar clause relating to such an unfortunate turn of events and the procedures to follow at any stage after the birth of the child until it reaches majority could also be inserted to assure the surrogate mother of the long-term care of the child.

The fact that the surrogate mother should "observe a sensible lifestyle, diet and standard of hygiene" throughout the pregnancy was viewed as a contractual stipulation which should be required by legislation.[202] Since the CA does not specifically require the inclusion of such an undertaking it is advisable that parties include it as *incidentalia*. How non-compliance with requirements of such a personal nature or those regarding sexual abstinence by the surrogate during the fertilisation period or partaking in exercise during the pregnancy could be enforced remains to be seen.[203] It is nevertheless advisable that all parties be made aware of the others' expectations in writing so as to avoid uncertainty and confusion.

The financial arrangements between the parties should be very clearly set out in order to protect all parties to the agreement. A surrogate mother should not suffer financial loss due to her pregnancy, nor should the commissioning parents be extorted to provide more than was agreed upon.[204] These financial clauses should specify all aspects that will be compensated for, as claims above and beyond what is stipulated might not be contractually enforceable. Amounts spent on maternity clothes should be identified and, where commissioning parents wish the surrogate mother to follow a specific diet, they may compensate her for groceries and dietary supplements if specifically and clearly defined and stipulated.

It seems preposterous that some of these optional provisions are not made mandatory by the CA. In light thereof it is expected of legal practitioners drafting surrogate motherhood agreements to foresee all possibilities that may arise during the period of the effective agreement. This surely is an impossible feat, and yet this is the practical implication

200 A maximum of three embryos per attempt is permitted, unless medical reasons determine otherwise (reg 12 *Regulations Relating to Artificial Fertilisation of Persons Supra* n 129); Nöthling Slabbert *SAJBL* June 2012 30.

201 S 295(d) CA.

202 Lupton 1986 *J Juridical Science* 155.

203 Pretorius 1996 *De Rebus* 114 115.

204 Unreported case. A surrogate from Bethlehem apparently demanded a motor vehicle and R100,000.00 before she would hand over the child. See *Beeld* (2011-01-22) 7; *Sunday Tribune* (2011-01-23) 3; *Daily News* (2011-01-26) 5; Nöthling Slabbert *SAJBL* June 2012 30. The incident even made headlines in the United Kingdom, see http://www.telegraph.co.uk/news/worldnews/africaandindianocean/southafrica/8269085/Surrogate-mother-holding-couple-ransom-with-unborn-baby.html (accessed 2012-09-12).

of poorly drafted legislation. A draft surrogate motherhood agreement has been formulated.[205] This was however done before the handing down of the *Ex parte WH* judgment. It serves as a good point of departure, but it should be remembered that it was drafted from a family law perspective and that it does not address all the requirements as set out by the courts in the most recent case law.

5 7 Breach

Stipulations relating to breach of contract by any of the parties should be addressed and illuminated in the *incidentalia*.[206] As surrogacy law and its adjudication in South African courts is still in its infancy, parties to surrogate motherhood agreements should act preventatively and include as much information in their agreements as possible. This will help the court in its attempt to balance the true intention of the parties with what Chapter 19 of the CA prescribes. The most obvious problem that could arise is that the surrogate mother, who is not genetically linked to the child, refuses to give up the child after its birth. Forcing her to do so has been described as "sacrifice[ing] a woman's reproductive autonomy to the principle of *pacta servanda sunt*".[207]

The Law Commission and the legislature endorse the application of the *pacta servanda sunt* principle.[208] Due to the nature of surrogacy, the only performance that will truly satisfy the aggrieved party in the case of breach is specific performance. No court will order damages, beyond actual expenses incurred, be paid to the commissioning parents by the surrogate mother, the result being that where a dispute arises there will never be justice for both parties.[209] The same applies to the situation where the surrogate mother wishes to undergo an abortion, to which she is legally entitled, and the commissioning parents wish to prevent this.[210] Louw[211] is of the opinion that compelling specific performance could possibly be unconstitutional as such an order might unduly impact the surrogate mother's rights to dignity, privacy and reproductive autonomy.[212] The extremely personal nature of the surrogate motherhood agreement has also been identified as a reason why specific performance would not be an appropriate order in the instance of breach.[213]

Lupton argues that prior to the CA, a court would have applied the *boni mores* test to determine whether a surrogate motherhood agreement would be enforceable.[214] It could be contended that even now that

205 Carnelley & Sheetal "Surrogate motherhood agreements" 2011 *De Rebus* 32.
206 Lupton 1986 *J Juridical Science* 155.
207 Clark 1993 *SALJ* 777.
208 *Idem* 773.
209 Meyerson 1994 *Acta Juridica* 121 141.
210 *Idem* 131; Louw 19-28.
211 At 19-23.
212 Respectively enshrined by ss 10, 14, 12(2) Constitution.
213 Lupton 1986 *J Juridical Science* 155, Clark 1993 *SALJ* 778.
214 1991 *TSAR* 229.

formal surrogacy is regulated by the CA, the court might still apply this test where either party to the agreement refuses to abide by it or to perform in terms of the contract. How South African courts will interpret section 298(1) in future remains to be seen. What will a court rule where in a case of full surrogacy the child is not related to the surrogate and she refuses to hand it over to the commissioning parent? Will the court in adjudicating custody disputes ascribe the highest emphasis to the parties' genetic link, or lack thereof, to the child? The question regarding the strict application of sections 298(1) and 297(1)(a) and (b) of the CA will only be answered if the Constitutional validity of these sections is tested.

6 Practical and Theoretical Approaches To Surrogacy

6 1 Practical and Procedural Aspects to be Kept in Mind by Practitioners

When drafting the surrogate motherhood agreement, attorneys are advised to refer to "the child(ren)" that is/are to be born as a result of the surrogacy, as artificial fertilisation often results in more than one child being conceived.[215]

If any previous applications for confirmation of surrogate motherhood agreements have been submitted by any of the parties all information regarding these applications should also be provided to the court in the affidavit.[216]

Directive 4.1 of Practice Directive of the South Gauteng High Court[217] stipulates that confirmation applications are not automatically urgent and that the Registrar of the court must issue them in the ordinary course. Thereafter, the entire file must immediately be taken to the office of the Deputy Judge President where the application will be assigned to a particular judge who may then further direct how the matter is to be heard and what further information should be provided to the court.[218] Counsel for the applicants should specifically direct the court's attention to the requirements contained in section 295 of the CA and indicate how they are met in the application.[219]

6 2 The Contract Law and the Family Law Models

Surrogacy stands astride public and private law since constitutional, administrative and criminal law principles on the one hand, and family, contractual and delictual law principles on the other hand, all apply to

215 Pretorius 1996 *De Rebus* 114 117.
216 *Ex parte WH Supra* 531D-E.
217 *Supra* n 70.
218 *Idem* directives 4.2, 4.3 & 4.5.
219 *Idem* directive 4.4.

this unique niche of the law.[220] Against the backdrop of childcare, child abduction, abortion, corporal punishment and surrogacy, Sloth Nielsen and Van Heerden discuss the need for a "paradigm shift" with regards to the doctrinal notion of the public law – private law divide.[221] This shift is of crucial importance in the context of surrogacy. When parties try to affect surrogacy amongst them, the debate further intensifies when the role of the private law approaches of contract law and family law are weighed up. Parties need to understand the role government bodies, the legislature and the courts play in the realm of the surrogate motherhood agreement. The notion of increased governmental intrusion into the private lives of citizens is undesirable in many instances; however, it must be welcomed in relation to surrogacy.[222]

One of the major policy issues in the legal approach to surrogacy is deciding on the appropriate paradigm to adopt in order to analyse surrogacy.[223] The two traditional, polar opposite options are a family relations or family law model, based on the principles of adoption, and that of contractual relations or the contract law model, based on the concepts of the sale of products or services.[224] Another, possibly more palatable, description of the two models is of the family law model as a "state regulation approach" and the contract law model as a "private ordering approach".[225] A detailed description of each model will not be provided, but the merits of the arguments for and against either will be evaluated.[226]

The CA implies[227] that disputes regarding the parentage of the child will be settled differently where the surrogate mother is also the genetic mother of the child. Where the surrogate is not the genetic mother, the family law model's stance is that the gestational mother (surrogate) has the strongest right to the child.[228] However, this comes into direct contrast with this model's approval of relying on the genetic link in cases of adoption where a genetic mother changes her mind about giving up her child.[229] Directly transplanting the rules of adoption to surrogacy (as the family law model advocates) is also not viable under South African law, due to the genetic link of at least one of the commissioning parents to the child. This automatically sets the surrogacy situation apart from that of adoption.

220 Pretorius 1996 *De Rebus* 114.
221 Sloth Nielsen & Van Heereden 1998 *SALJ* 159-164.
222 *Idem* 164.
223 Capron & Radin 1988 *Law, Medicine and Health Care* 34.
224 *Ibid*.
225 Pretorius 1996 *De Rebus* 114 115.
226 Most of the literature on the paradigms dates from the 1980s when surrogacy and artificial fertilisation were not as advanced as they are today and may thus not be appropriate in the current context.
227 In s 298(1) CA.
228 Capron & Radin 1988 *Law, Medicine and Health Care* 35.
229 *Ibid*.

However the invaluable role of certain aspects of the family law model is undeniable. The essential involvement of the State through courts, social service institutions and legislative regulation is closely related to the family law model, as is the State's responsibility as upper guardian of all children.[230] The family law model is preferred by Capron and Rodin, but no discussion of the merits of the contract law model is considered.[231] Their explanation of the family law model involves that the child be adopted, the possibility of which is not clear under the CA.[232]

The mere fact that a document to which everyone involved is a party is aimed at altering the effect of the existing common law as it is related to the family must mean that the role of the law of contract cannot be ignored. A rejection of the contract law model inadvertently means the rejection of the rights and duties associated with the law of contract, namely freedom of contract and *pacta servanda sunt*. Pretorius favours the contract law model, stating that it can successfully regulate surrogacy.[233]

The contract law model has been discarded in the United States of America due to the doctrine of valuable consideration which applies under American law.[234] An automatic rejection of the contract law model in South Africa cannot be based on the same logic, as the doctrine is not applicable in South African law.[235] Therefore the argument that promises without consideration (such as altruistic surrogacy) are not enforceable, cannot apply in the South African context.

The contract law model cannot be described as perfect either. If the parties to the agreement's relationship to the child rests on the validity and accuracy of a document like a surrogate motherhood agreement, and the validity of the document or its actual intention comes into question, the child's care and parentage is automatically jeopardised.[236] Moreover the contract law concepts related to a transaction where "ownership" in a thing is being transferred from one individual to another[237] is also highly unacceptable to the South African embodiment of the value of a human being. The traditional contracts of sale or letting and hiring of services remind too much of warranties, guarantees and the enforcement of performance by any legal means necessary. Consumer law and the effects thereof can under no circumstances be applied to the law of surrogacy.

Meyerson correctly argues that the situation regarding surrogacy is more complicated than allows it to squarely fit into the box of either the

230 *Idem* 37.
231 *Idem* 34.
232 *Supra* n 120.
233 1996 *De Rebus* 115 115-116.
234 Capron & Radin 1988 *Law, Medicine and Health Care* 40.
235 *Conradie v Rossouw* 1919 AD 279 298.
236 Capron & Radin 1988 *Law, Medicine and Health Care* 35.
237 *Idem* 36.

family law or contract law model: surrogacy is more complex than adoption and the surrogate motherhood agreement cannot be seen as just an ordinary contract.[238] Because of the "unique mixture of contractual and personal interests" choosing one model over the other would not allow for dealing with the situation as effectively or temperately as possible.[239] A mixed model, encasing aspects of both models should be designed.[240]

Furthermore, the fact that the general principles of the law of contract alone were deemed insufficient to regulate surrogacy means that a blending of the contract law and family models was what the Law Commission intended.[241] Academic research should thus be conducted on how to devise a surrogacy specific model suited to the current South African legal climate.

7 Conclusion

Despite the fact that the recent case law has cleared up some uncertainties regarding the law of surrogacy, many legal and ethical dilemmas remain.[242] Despite current *lacunae* the contract serves as a tool to clarify and protect the interests of both the commissioning parent(s) and the surrogate mother.

The courts demand a complete, honest and detailed contract, affidavit and application. Despite this, parties seem to either not know, or care about the effect of their non-compliance. Before surrogacy was specifically legislated by the CA, it is believed that the practice was driven underground and still practiced by infertile couples.[243] The fact that surrogacy is now legislated for begs the question, why do parties persist in flouting the law?

Surrogacy is inevitably engaged in the intricate web of trying to balance the interests of the various parties to the agreement, the child and society as a whole – a seemingly impossible feat.[244] At present the surrogate motherhood agreement is the best tool available to the parties to attempt to clarify and balance these individual interests. The possibility of a surrogacy specific model should be researched in order to provide prospective parties to surrogate motherhood agreements, as well as our courts, guidance in interpreting, enforcing and implementing surrogacy agreements under Chapter 19 of the CA.

238 Meyerson 1994 *Acta Juridica* 121 144.
239 *Ibid*.
240 *Ibid*.
241 *Supra* n 66.
242 Nöthling Slabbert *SAJBL* June 2012 31.
243 Pretorius 1991 *De Jure* 52 61. Lupton 1986 *J Juridical Science* 157 states that when the practice of surrogacy is driven underground, the possibility of exploiting surrogate mothers becomes significantly higher.
244 Lupton 1986 *J Juridical Science* 156; Nöthling Slabbert *SAJBL* June 2012 31.

Until such time as a surrogacy specific model has been designed, or a court hands down a judgment implying otherwise, it is suggested that the contract law model be adopted. It provides the best perspective from which to give effect to the parties' intentions. Furthermore, it is submitted that the reigning uncertainties resulting from the lack of regulations to Chapter 19 of the CA are better remedied by applying the contract law model.

Dawn of a competitive electricity sector for South Africa: The Independent System and Market Operator Bill B 9-2012 – Context, content and comment

Elsa Steyn
BA LLB Cert Adv Corp Law
Advocate of the High Court of South Africa

OPSOMMING

Aanvang van 'n Mededingende Elektrisiteits-sektor in Suid-Afrika: *The Independent System and Market Operator Bill B9-2012* – Konteks, Inhoud en Kommentaar

Die *Independent System and Market Operator Bill B9-2012* ("ISMO-Wetsontwerp") wat tans voor die Parlement dien, is 'n kardinale stap in die ontwikkeling van mededinging in die Suid-Afrikaanse elektrisiteitsektor. Die Wetsontwerp word hier krities beskou in die konteks van mededingingsverwikkelinge in die sektor en enkele vergelykings word getref met die Ontario Independent Electricity System Operator (IESO) ten einde voorstelle ter verstewiging van die Wetsontwerp aan die hand te doen. Hierdie het hoofsaaklik betrekking op die korporatiewe vorm van ISMO; ISMO se drievoudige rol met betrekking tot die elektrisiteitbeplanningsfunksie, stelselontwikkeling en as stelsel- en markoperateur; en die eienaarskap van die transmissiebates. 'n Deurtastende beskouing van mededinging in die elektrisiteitsektor toon aan dat velerlei ondernemings in dié verband gemaak is deur die owerhede sedert 1998, maar dat daar weinig van gekom het tot dusver, met die resulterende kapasiteitskrisis wat weer in die winter van 2012 gelei het tot kragonderbrekings en beurtkrag. Daar word voorgestel dat ISMO 'n groter rol te speel het as wat tans in die Wetsontwerp beoog word in verband met elektrisiteitsbeplanning, die toewysing van nuwe opwekkingsgeleenthede, die ontwikkeling en uitbreiding van die nasionale transmissiestelsel en die kontraktering vir nuwe opwekking. Dit word ook beklemtoon dat die eienaarskap van die transmissiebates 'n deurslaggewende faktor kan wees in die uiteindelike sukses aldan nie van ISMO. Die belang van 'n onafhanklike, neutrale mark- en stelseloperateur kan nie oorbeklemtoon word met verwysing na die Renewable Energy Independent Producer Procurement Programme wat einde 2011 deur die Departement van Energie van stapel gestuur is nie. Dit is 'n voorvereiste vir 'n suksesvolle belegging deur 'n onafhanklike kragprodusent dat hy onbelemmerde toegang tot die transmissienetwerk moet kry en dat hy van 'n afset verseker is vir sy produk. ISMO vervul juis hierdie rol.

1 Introduction

South Africa has a vertically integrated power utility in the form of Eskom Holdings Ltd ("Eskom"). It is for all intents and purposes a total

monopoly, controlling generation,[1] transmission[2] and distribution[3] of electricity throughout the Republic. Generation, transmission and distribution represent the three levels at which competition might present itself, given an enabling regulatory framework. Unbundling and deregulation of the electricity market thus typically involves the "splitting up" of utilities such as Eskom into separate Generation, Transmission and Distribution companies and the introduction of an Independent System Operator entity that functions autonomously from these companies, typically to operate the national grid and to fulfil a planning and procurement function in respect of new generation capacity. This ensures that in a system where a generator (upstream) such as Eskom also owns transmission assets such as the national grid and distributes electricity to retail customers (downstream), new private sector participants in the generation sector are ensured fair access to the national grid as well as an offtaker for the electricity generated by them. Without guarantees of equal access to the national grid and offtake for electricity generated, Independent Power Producers (IPPs) would find it close to impossible to have a bankable project for development.

Access to the national grid could be frustrated in several ways by the generator-owner: It could refuse the IPP a connection to the grid; it could delay the IPP's connection in favour of its own new developments; or it could charge the IPP prohibitive connection and/or wheeling charges – the latter refers to a charge for the conveyance of electricity over the grid for offtake elsewhere by a customer. Thus, for instance, if the IPP is a wind farm generator and its customer is an industrial steel plant some 100 km away, the IPP would upload electricity onto the national grid at point A, "wheel it" over the grid and the customer would download an equivalent quantity of electricity from the national grid at point B. The offtake issue arises by virtue of the fact that the generator-owner of the grid typically has lower sunk costs than the IPP by reason of its existing infrastructure (it can finance new developments with cross-subsidisations from existing generation facilities) and can afford to sell electricity at lower tariffs than the IPP, meaning that the playing ground is not level. It is clear, therefore, that an independent system operator is pretty much a prerequisite for the introduction of competition into an electricity market, preferably accompanied by the unbundling of the incumbent vertically integrated monopoly utility.

Although the Government indicated its intention to introduce competition in the electricity sector as long ago as 1998,[4] it was only in 2006 that the framework for private sector participation in electricity generation was created with the enactment of the Electricity Regulation

1 The production of electricity by any means.
2 The conveyance of electricity through a power network above a nominal voltage level of 132 kV.
3 The conveyance of electricity through a power network at or below a nominal voltage level of 132 kV.
4 See s 2 below.

Act[5] (ERA). Yet, despite the provisions of the ERA which enable the Minister of Energy to promulgate regulations pertaining to new generation capacity, private sector participation, the procurement process and the cost recovery mechanism to be utilised, the present monopolistic structure of the electricity sector still does not provide for the independent purchase of power from the private sector.[6] That is due to first, the fact that unbundling of Eskom into different generation, transmission and distribution entities has not been seen through, and second, the absence of an independent system operator entity in the South African electricity market. It is a major stumbling block to the introduction of competitive private sector participation in the South African power sector: predictably, barriers such as impeded access to the transmission system, delayed connections to the transmission system and unequal bargaining power for the sale of electricity where the buyer also generates, arise.[7]

The Department of Energy (DoE) has recognised that a conflict of interest in the current regulatory framework is inevitable for the reasons cited above[8] and has accordingly put forward the Independent System and Market Operator Bill[9] ("the ISMO Bill") that was introduced in the National Assembly as a proposed section 75 bill in March 2012. The stated intention with the ISMO Bill is to create a market and system operator that will be independent from activities related to electricity generation, to avoid a conflict of interest, and thereby to ensure the equal treatment of all generators.[10]

The concept of an independent system operator has never been more topical, as the DoE has already embarked on an extensive procurement drive for IPPs in the renewable energy generation sector, with the Minister of Energy declaring her intention to procure 3,725 MW of investment in renewable energy projects from the private sector. She has so far awarded projects amounting to 2,455 MW in the first two rounds of bidding. To date some R50 billion has reportedly been pledged by the private sector in what has been described as an "energy success story that could spark a chain reaction across Africa."[11]

---

5 4 of 2006.

6 See par 1.2 *Memorandum on the Objects of the Independent System and Market Operator Bill, 2012*.

7 In a Departmental Briefing to the Parliamentary Portfolio Committee on 2011-07-07, the DoE ascribed the fact that IPPs had until then not been forthcoming in significant volumes, to four factors: "[The p]erception of conflict of interests in vertically integrated Eskom; [the p]erception that Government is not serious about industry reform; [p]erceptions about long-term viability of present ESI structure; [and the l]ack of enabling legal/ regulatory framework to facilitate IPPs."

8 Departmental Briefing to Parliamentary Portfolio Committee 2011-07-07

9 B 9-2012.

10 Par 3.1 *Memorandum on the Objects of the Independent System and Market Operator Bill, 2012*.

11 Nott "Werksmans Africa Legal Brief Series" March 2012.

This article will place the development of the ISMO Bill in historical context, consider the content of the ISMO Bill, and suggest improvements against the background of the experience of a similar entity in North America. Due to its limited scope it is not intended to be exhaustive.

2 Context – Competition and the South African Electricity Sector

Modern-day electricity industry regulation in South Africa commenced with the establishment of the National Electricity Regulator (NER) in 1995 following the publication of an amendment to the Electricity Act[12] by the Electricity Amendment Act,[13] with the objective of introducing efficient and effective regulation of the Electricity Supply Industry (ESI) in line with Government policy and law. This entity continued in existence until 2006, when it was replaced by the National Energy Regulator of South Africa (NERSA),[14] the body that remains responsible for electricity regulation to this day.

The first steps towards encouragement of competition in the ESI are to be found in the *White Paper on the Energy Policy of the Republic of South Africa*, gazetted in 1998:[15] two of the policy objectives of the White Paper are (1) "stimulating economic development through the encouragement of competition in energy markets, encouragement of cost-reflective tariffs and creation of investor-friendly climate"; and (2) "securing supply through diversity, given increased opportunities for energy trade."[16] The White Paper furthermore prescribes that the Integrated Resource Planning (IRP) methodology is to be followed for energy planning purposes[17] – this would play a big role later on, as it essentially prescribes what is permissible in terms of new capacity to be sourced, both in terms of energy diversity and private sector participation. It provides for the introduction of IPPs into the generation market;[18] the restructuring of Eskom through the unbundling of its generation and transmission groups;[19] and for improved electricity sector governance, *inter alia* to introduce and maintain effective competition within the electricity industry.[20]

In April 2001 Cabinet took a formal decision to restructure the ESI. The strategy included three important aspects for present purposes: (1) the

12 41 of 1987.
13 46 of 1994.
14 In terms of the National Energy Regulator Act, 40 of 2004 read with GN 666 GG 28996 of 2006-07-07.
15 GG 19606 of 1998-12-17 (the "White Paper").
16 *Idem* 6.
17 *Idem* 18.
18 *Idem* 20.
19 *Idem* 20.
20 *Idem* 21.

unbundling of Eskom; (2) the selling of 30 % of electricity generation; and (3) gradually increasing private sector participation by ensuring that all new generation capacity is built by the private sector.[21]

This was in line with Government's goals for State Owned Enterprises (SOEs) at the time, which included: the mobilisation of private sector capital and foreign direct investment; enhanced competitiveness of SOEs; the promotion of fair competition; and the financing of growth and requirements for competitiveness.[22] The NER accordingly advocated a three-phased approach to reform to a competitive ESI sector: first, Eskom corporatisation; second, the corporatisation of generation and independent transmission; and third, private sector involvement and competition.[23]

Eskom's own proposals at the time are instructive. It warned that a cautionary approach had to be followed as the implications of restructuring errors would be severe and difficult to reverse.[24] Therefore it also foresaw a phased process, in terms of which over the course of 2001-2002 Eskom would be converted into a company and a new board of directors appointed. Generation, transmission and Distribution would be set up as operating divisions of the company, with the option of establishing the three entities as subsidiaries of Eskom to be investigated. The period 2002-2003 would then see the possible incorporation of Generation, Transmission and Distribution as wholly-owned subsidiaries of Eskom (in line with the investigation to be performed). Generation restructuring would be conducted against the principle that Eskom would progressively reduce its local market share whilst expanding internationally: 2001-2002 would see the introduction of internal competition (clusters) and the finalisation of planning for Black Economic Empowerment (BEE) and private participation and/or the gradual sale of some generation assets; and 2003-2006 would see the implementation of plans for BEE and private participation and/or the gradual sale of some generation assets.[25]

Transmission restructuring would first entail the ringfencing of Transmission into two entities, a wireless/system operator business and a market operator (power exchange) (2001-2002); next Eskom would consider establishing Transmission as a subsidiary to establish independence and create "Chinese Walls" to avoid Generation being able to influence the market operator (2002-2003); and ultimately a separate power exchange would be created from Eskom and set up as an independent entity in synchronisation with the gradual introduction of a

21 *Energy Security Master Plan – Electricity 2007-2025* (2007) 7 (the "Energy Master Plan").
22 NER briefing to Standing Parliamentary Committee on Minerals & Energy "On ESI and EDI Restructuring" 2001-09-19.
23 *Ibid*.
24 Eskom briefing to Standing Parliamentary Committee on Minerals & Energy "Eskom Strategic Intent and Restructuring" 2001-10-09.
25 *Ibid*.

market and the sale of some Generation assets (subject to regulatory and market rules being in place), the transmission wires and system operator business remaining with Eskom and longer-term positioning of Transmission to be determined (2003).[26]

Distribution restructuring would be conducted largely in line with the Electricity Distribution Industry (EDI) restructuring process of DME, with Eskom retaining some wholesale and retail customer services capability.[27]

Eskom was indeed converted into a public company with effect from 1 July 2002 and it created the three divisions as detailed above.[28] However, nothing came of the mooted potential incorporation of the various divisions into separate subsidiaries and thus the introduction of "Chinese walls". When faced with the impending reality of an independent system operator entity in the South African market, Eskom responded by creating such an entity within one of its subsidiaries, Eskom Enterprises Ltd, and it has consistently argued against the incorporation of a new entity as provided for in the ISMO Bill, proposing instead that the Eskom unit should be utilised.[29] That creates obvious problems of independence and neutrality, as will be discussed in further detail below.

In 2003 Cabinet approved private-sector participation in the electricity industry and resolved that future power generation capacity would be divided between Eskom (70 %) and the IPPs (30 %). The then Department of Minerals and Energy (DME) was mandated to ensure private-sector participation in power generation through a competitive bidding process that would diversify primary energy sources and be developed within the electricity sector.[30] However, it would take until the promulgation of the ERA in 2006 before any legislative framework for the procurement of generation from IPPs would come into existence.

The first mention of an Independent System Operator (ISO) is to be found in the draft *Electricity Pricing Policy of the South African Electricity Supply Industry*, gazetted for public comment in May 2004.[31] It proposed the ESI to operate on the principles of a "multi-market model with a number of generation clusters, IPPs, and independent Transmission Company and a Transmission Systems Operator."[32] This draft policy was never adopted.

---

26 *Ibid*.
27 *Ibid*.
28 S 3(1) Eskom Conversion Act 13 of 2001, read with Proc 56 GG 23561 of 2002-06-28.
29 Eskom Presentation to the Portfolio Committee on Energy "Eskom's Comments on the ISMO Bill" 2012-05-16.
30 *2007 National Energy Summit* hosted by DME Notice 504 GG 32230 of (20070925–27) 53 (the "Energy Summit").
31 GG 2004-05-28.
32 *Idem* 10.

In 2006 the Electricity Act[33] was replaced by the ERA. It was the culmination of nearly two years' worth of draft bills, sometimes vigorously contested. One such area of disagreement pertained to the introduction of private sector generation: Cosatu was strongly opposed to this and argued that the proposed section 46 dealing with it in the draft Bill[34] should be scrapped in its entirety. The DME rejected this and commented that "[p]rivate participation in generation is policy."[35] This was confirmed by the Minister of Minerals and Energy, Lindiwe Hendricks, in her submission to the Parliamentary Portfolio Committee when she cited the policy requirements of the *White Paper* for private sector participation, saying:

> The Bill provides for non-discriminatory access to the electricity networks. This is to ensure participation of IPPs, including renewable energy producers. Government's strategy to create jobs includes the facilitation of Foreign Direct Investment (FDI) and Black Economic Empowerment (BEE). The introduction of IPPs has been identified as an opportunity that is in line with that strategy within the electricity sector.[36]

The process envisaged in section 34[37] of the ERA calls for a Ministerial determination that additional generation capacity is required to ensure the continued uninterrupted supply of electricity,[38] in terms of which the Minister may also require that new generation capacity must be established through a tendering procedure which is fair, equitable, transparent, competitive and cost-effective;[39] and which may provide for private sector generation.[40] This apparently open-ended approach was somewhat tempered by a 2007 Cabinet decision in terms of which Eskom was to be designated as the single buyer of power from IPPs in South Africa – a policy that was motivated with reference to "government's position that the security of supply was a national priority which took precedence over all other key elements, including a competitive market."[41]

The shift in Government policy was confirmed with the publication of the *Energy Security Master Plan – Electricity 2007-2025* by the DME later that year. With reference to the objectives of introduction of competition in the generation sector, the permission of open, non-discriminatory access to the transmission system and the encouragement of private sector participation in the electricity industry, the Master Plan states that

33 41 of 1987.
34 B 25 of 2005.
35 Parliamentary Portfolio Committee "Comments on the Electricity Regulation Bill" (2005-11-05) 6 1920.
36 Presentation by Minister of Minerals & Energy Lindiwe Hendricks to the Parliamentary Portfolio Committee on "Electricity Regulation Bill" (2005-11-07) 3.
37 Originally s 46, renumbered as s 34 by the *Electricity Regulation Amendment Act*, 28 of 2007.
38 S 34(1)(a).
39 S 46(1)(e)(i).
40 S 46(1)(e)(ii).
41 *Energy Summit* 55.

various international events had impacted significantly on electricity markets worldwide[42] and that this meant that a different strategy to the planning and procurement of new generation capacity had to be adopted – the *raison d'être* of the *Master Plan* itself. In terms of the *Master Plan* Eskom would continue to provide 70% of new capacity and IPPs would provide the balance, confirming the 70%-30% Cabinet policy decision of 2003.[43] It notes that there is:

> [a] very weak case for full competition on a merchant basis, ie it is anticipated that any private participation in the electricity industry will be via the IPP mechanism with a power purchase agreement with Eskom (single buyer model)[44]

thus also giving effect to the 2007 Cabinet decision.

The next mention of the System Operator is to be found in the Draft Regulations on Electricity Regulation published by the Minister of Minerals and Energy for public comment in terms of the ERA in February 2009:[45] these were ultimately absorbed into the final Electricity Regulations on New Generation Capacity (ERNGC) gazetted in August of 2009.[46]

The ERNGC applied to all types of generation technology including renewable generation and cogeneration, but excluding nuclear generation technology[47] and the objectives included the regulation of entry by a buyer and an IPP into a power purchase agreement and the facilitation of the fair treatment and the non-discrimination between IPP generators and the buyer.[48] The System Operator was to be responsible for the entire development process of the Integrated Resource Plan (IRP), except for its approval and gazetting, which was to be the responsibility of the DoE. This it would have to conduct in consultation with the Department and the Regulator.[49] The Regulations furthermore provided that the System Operator would primarily be responsible for procurement.[50] As there was no System Operator in place at this time, these provisions were largely academic in ambit.

Despite the fact that President Jacob Zuma announced in his State of the Nation address on 11 February 2010 that an Independent System and Market Operator would be established, the ERNGC were replaced by a new set of similarly named regulations in May 2011[51] from which all

42 Said to include, *inter alia*, "the collapse of energy markets in Ontario in 2002" (*Energy Master Plan* 7). As appears from 4 1 below, the Ontario energy markets have actually thrived since 2002.
43 11-12, 56.
44 55.
45 Notice 139 GG 31891 of 2009-02-13.
46 GN R 721 GG 32378 of 2009-08-05.
47 Reg 1.
48 Reg 2.
49 Reg 3(1), (2).
50 Regs 6, 7.
51 GN R 399 GG 34262 of 2011-05-04.

mention of System Operator had been expunged: the development of the IRP is now the responsibility of the Minister of Energy after consultation with the Regulator[52] and the procurement process no longer provides for processes conducted by the System Operator but rather for an "IPP procurement programme" which shall take the form determined by the "procurer", who may also be the buyer thereof.[53] The concept of "procurer" is not defined in the new regulations.

Rather incongruously, the next document to be gazetted by the Department of Energy was the publication of the draft ISMO Bill exactly nine days later on 13 May 2011.[54]

Final developments included the gazetting of two other draft Bills for public comment by the DoE in December 2011: the draft National Energy Regulator Amendment Bill[55] and the draft Second Electricity Regulation Amendment Bill,[56] as well as the introduction of the revised ISMO Bill in the National Assembly in March 2012.[57] These three Bills are being treated by the DoE as something of a triad, with the text of the revised ISMO Bill being drafted to reflect references to the ERA as proposed to be amended in the draft Second Electricity Regulation Amendment Bill, rather than the ERA in its present format.

3 The Independent System and Market Operator Bill

The stated object of the Bill is to provide for an ISMO as a company that is financially viable and responsible for the planning of supply of electricity by generators through the national transmission system, for electricity dispatch and aggregation in respect of sale of electricity by generators, to act as the buyer of electricity from generators and to sell electricity to ISMO customers.[58]

The structure of ISMO as proposed is contentious. The Bill foresees the incorporation of a for-profit company with the State as its sole shareholder.[59] Eskom has argued that the Bill should either allow for the ISMO to be established through the take-over by the State of the subsidiary created by Eskom for that purpose in Eskom Enterprises, or alternatively, it should allow for the transfer of the subsidiary from Eskom to a company established by the State.[60] Organisations such as

52 Reg 4(1).
53 Reg 7.
54 Notice 290 GG 34289 of 2011-05-13.
55 Notice 890 GG 34825 of 2011-12-08.
56 Notice 905 GG 34870 of 2011-12-19.
57 B9-2012.
58 S 2.
59 S 3.
60 Eskom Presentation to the Portfolio Committee on Energy "Eskom's Comments on the ISMO Bill" 2012-05-16.

Earthlife Africa, SAFCEI and NUMSA have taken issue with ISMO profit-making, essentially arguing that it would cause unnecessary increases in the electricity tariff and cause ISMO to make a profit in the performance of a public good.[61]

ISMO's functions are threefold: planning, system operation and expansion, and market operation. However, much criticism has been levied against the Bill by reason of the fact that the planning function is limited to the modelling of scenarios for purposes of developing the IRP and the provision of inputs to the transmitter for the development of the expansion plan in accordance with anticipated electricity demand as per the IRP. It has been argued that the Minister of Energy, and indeed the DoE, does not possess the necessary manpower and resources for the compilation of the IRP and that this should instead be the responsibility of ISMO itself as it will have both the data and resources at hand.[62] It is worth pointing out that the development of the IRP 2010,[63] which is currently in force, was largely conducted by Eskom at the behest of the DoE. The IRP itself is of cardinal importance for purposes of the present discussion, as a Ministerial determination in terms of section 34 of the ERA is required that new generation capacity is needed before an IPP procurement process may be embarked upon – and it is intended that in making such a determination the Minister will need to have regard to the content of the IRP.[64] It contains strict prescriptions regarding the energy mix and energy sources to be employed for new generation – and indeed, the section 34 determination must include provisions dealing with *inter alia* the types of energy sources from which electricity must be generated, and the percentages of electricity that must be generated from such sources.[65]

As market operator, ISMO is responsible for the procurement function and acts as buyer of electricity generated. The intention of DoE is to "level the playing field" so that Eskom does not act as both buyer and generator.[66] This would appear to tie in with the single buyer-model thus far mooted by Government, but industry players have expressed the

61 See Taylor "Sustainable Energy & Climate Change Project: Comments on the draft Independent System and Operator Bill – B9-2012" (2012-04-13); McDaid "SAFCEI Response to Publication of Draft Legislation – Independent System and Market Operator Bill – B9-2012" (2012-04-12); NUMSA "Submission Independent System & Market Operator Bill (ISMO Bill) – Additional Comments) (2012-05-23).

62 See Eberhard & Kapika "Independent System and Market Operator Bill – Comments to Parliamentary Committee on Energy" (2012-05-19); Pickering "Meridian Economics: Comments on the Independent System and Market Operator Establishment Bill" (2012-04-13) 3-5.

63 *Integrated Resource Plan for Electricity 2010-2030* Rev 2 Final Report (2011-03-25).

64 S 34(5) Electricity Regulation Second Amendment Bill. One concern with the Renewable Energy Procurement Programme is that no such announcement has been made to date in respect of the 3,275 MW that is apparently being procured from IPPs.

65 S 34(1)(b) ERA.

66 Departmental Briefing to Parliamentary Portfolio Committee (2011-07-07).

hope that there would be scope for bilateral agreements[67] in view of the fact that the Electricity Regulation Second Amendment Bill provides a Ministerial exemption from the single buyer obligation[68] and the ISMO Bill itself refers to "generation licensees exempted by licence from selling to ISMO in terms of the Electricity Regulation Act".[69] There would accordingly appear to be the potential for ISMO as a non-exclusive central purchaser, which means that some competition in the wholesale power market may be possible, with attendant efficiencies to be gained.

In this regard, there is a lack of clarity regarding the identity of ISMO customers.[70] The Bill defines it broadly as "the customers to whom ISMO may sell electricity; from whom ISMO may buy electricity; and who provide ancillary services." Predictably, submissions to the Parliamentary Portfolio Committee reflect the underlying market philosophy of the commentator in question, ranging from NUMSA who believes that all customers should be ISMO customers,[71] to the Free Market Foundation, who believes that there should be free competition,[72] to the extent that "government should pave the way for the entry of competing suppliers of every possible service involved in the process, from generation to distribution".[73]

With regards to the national grid, it is unclear from the Bill whether ISMO will own or will simply operate the transmission network system. The Bill provides for a procedure whereby Eskom must compile a list of all assets, rights, liabilities and obligations to be transferred to ISMO, together with the value thereof.[74] This includes a provision for the transfer of employees who perform functions pertaining to ISMO.[75] It is not clear what the extent of these assets is, but arguably it could include the transmission system. This is an issue of great importance, as it determines to a large extent the balance sheet of ISMO – which will

67 Kruger "Chamber of Mines of South Africa: Comment on Independent System and Market Operator Bill – B9 2012" (2012-04-16); Pickering *op cit*; Energy Intensive Users Group (EIUG) "Independent System and Market Operator Establishment Bill [B8-2012]: Parliamentary Submission" (2012-04-18); Eberhard & Kapika *op cit*.

68 See s 34(2)(c).

69 S 4(3)(d)(i). Note that the ISMO Bill is written throughout as though the Electricity Regulation Second Amendment Bill has already been enacted – the present ERA does not provide for such exemption by license.

70 See eg SALGA "Presentation to the Portfolio Committee on Energy: Independent Systems and Market Operator Bill Public Hearings" 15 and 16 May (2012-05-15–16).

71 See NUMSA *op cit*.

72 Free Market Foundation "ISMO Bill Submission" (2012-05-22).

73 Free Market Foundation "Comment on the ISMO Draft Bill" available at http://eepublishers.co.za/article/free-market-foundation-103.html (accessed 2012-05-24).

74 S 40. NERSA has pointed out that Eskom's current transmission licence will need to be amended in order to separate the system operator part: "NERSA Presentation to Portfolio Committee on Energy" (2012-05-16).

75 See in this regard Eskom Presentation to the Parliamentary Portfolio Committee *op cit*; NUMSA *op cit*; and Nedlac "Nedlac Report to the National Assembly Portfolio Committee on Energy on the ISMO Bill" (2012-05-22).

impact on the bankability of power purchase contracts entered into with IPPs. The PPAs would otherwise require a government guarantee to support the agreement and make it bankable.[76] Not having a bankable project in this sense then becomes a further barrier to IPP generation and thus the introduction of competitive private sector participation in the South African electricity market. Eberhard and Kapika also argue that the transfer of the transmission assets to ISMO would create a level playing field between Eskom and the IPPs[77] and point out that the argument that this will ruin Eskom's credit rating is fallacious, since the value of Eskom transmission assets is minute compared to the remainder.[78] This is borne out by a referral to Eskom's 2012 Integrated Report, in terms of which the carrying value of Transmission Assets amounts to R15.8 billion of the total Property, Plant and Equipment portfolio of R290.6 billion – some 3.8%.[79]

Finally, there is a provision for ISMO to contract with Eskom for the execution of ISMO functions, presumably to ensure the seamless transition between the *status quo* and the implementation of ISMO. It is unclear from the wording of the clause whether this is intended to be just an interim measure until ISMO is up and running or whether Eskom may continue to do so indefinitely. It goes without saying that there are grave objections to the latter scenario,[80] as the very *raison d'être* of ISMO is to create an entity that is "independent from all activities related to electricity generation to ensure equal treatment of all generators."[81]

4 Comparison: Ontario Independent Electricity System Operator

4 1 Background

The Ontario Independent Electricity System Operator (IESO) provides an interesting basis for comparison. As is the situation with Eskom in Southern Africa,[82] Ontario is a net exporter of electricity in North America – in 2011, for example, it exported a net balance of 8.9 TWh of electricity, to a value of CAD 277.1 million.[83]

76 Free Market Foundation *op cit*.
77 Also see Kruger *op cit*.
78 *Ibid*.
79 Statement of Financial Position available at http://financialresults.co.za/2012/eskom_ar2012/ financial-statements/fin-position.php (accessed 2012-07-23).
80 See eg EIUG *op cit* 8-9.
81 Par 2.3 *Memorandum on the Objects of the Independent System and Market Operator Bill*, 2012.
82 Eskom generates approximately 95% of the electricity used in South Africa and 45% of the electricity used in Africa: see http://www.eskom.co.za/c/40/company-information/ (accessed 2012-07-23).
83 See http://www.ieso.ca/imoweb/siteShared/imports_exports.asp?sid=ic (accessed 2012-06-22).

In 1998, for purposes of creating a competitive electricity market, the former Ontario Hydro's monopoly in the electricity market was replaced with several business entities, including two distinct commercial companies: Ontario Power Generation Inc. (OPG)[84] and Ontario Hydro Services Company Inc. (Hydro One),[85] as well as one Crown corporation, the Independent Electricity Market Operator (now IESO).[86] OPG is responsible for the generation of electricity, while Hydro One owns and maintains transmission and distribution wires.[87] IESO manages the province's power system and is responsible for the wholesale electricity market. The Ontario Energy Board (OEB) regulates the province's electricity and gas sectors and has varying degrees of regulatory authority over the three entities, as well as the province's municipal electric utilities.

These structural changes were the precursor to the opening of the Ontario electricity market on 1 May 2002, which functions as a completely open but regulated competitive market.

December 2004 saw the creation of a new agency, the Ontario Power Authority (OPA), to ensure an adequate, reliable and secure supply of electricity in Ontario for the medium and long term.[88] The OPA is licensed by the OEB and the OEB approves its fees and its Integrated Power System Plan and procurement process.

4 2 Corporate Entity

The IESO is a not-for-profit company vested in the Crown but is not a Crown agency company.[89]

For South African purposes, it would appear correct to incorporate ISMO as an SOE (profit company) and to do so *de novo*, despite Eskom's objections in this regard.[90] Although there is much to be said for the argument that ISMO should be a not-for-profit company as is the case with IESO, the structure of SOE's in South Africa demands that it be treated congruently.[91] It appears dubious, however, whether ISMO could as an SOE fall under the jurisdiction of the Minister of Energy rather than the Minister of Public Enterprises, as envisaged by the Bill.[92] The argument against talking over an existing division of Eskom is one of

84 Incorporated under the Business Corporations Act, 1990 on 1998-12-01.
85 *Ibid*.
86 Ie corporation without a share capital: s 4(1) Electricity Act, 1998 as amended by s 23 Sched A s 4(1) Electricity Restructuring Act, 2004.
87 See s 48(1) Electricity Act.
88 The OPA is a not-for-profit Crown corporation without share capital, established in terms of s 25.1(1) Electricity Act.
89 S 5(2), read with s 5(4), (6). It is also expressly excluded from the ambit of the Business Corporations Act, 1990: s 24 Electricity Act.
90 See s 3 and *supra* n 58.
91 NUMSA *op cit* has argued that SOEs need to be "resocialised" by "change[ing] their mandates from the current profit orientation to service provision." That would provide one way out of the current conundrum.
92 Also see McDaid *op cit*.

neutrality, which is core to the function that ISMO is to fulfil in the market.

4 3 Objectives

IESO's statutory objectives include entering into agreements with transmitters giving IESO authority to direct the operation of their transmission systems; directing the operation and maintaining the reliability of the IESO-controlled grid; participating in the development of standards and criteria relating to the reliability of transmission systems; working with the responsible authorities outside Ontario to co-ordinate the IESO's activities with their activities; collecting and providing to the OPA and the public information relating to the current and short term electricity needs of Ontario and the adequacy and reliability of the integrated power system to meet those needs; and operating the IESO-administered markets.[93] It is clear from the above that the ambit of ISMO's objectives goes wider than those of IESO. The objectives of OPA are therefore also considered for purposes of this discussion.

4 4 Functions

4 4 1 Planning

It is important to note that both the planning function and the procurement function in Ontario vest in the OPA and not in IESO.[94] This includes the development of an Integrated Power System Plan on a periodical basis that will enable the Ontario Government to achieve its goals relating to the adequacy and reliability of electricity supply.[95] It is also responsible for a periodical assessment of the adequacy and reliability of electricity resources with reference to anticipated electricity supply, capacity, reliability and demand for the period in question, for purposes of which it must consider both generation and transmission capacities and technologies, as well as conservation measures.[96]

93 S 5(1).

94 OPA may, subject to OEB approval, set fees and charges to recover the cost of anything that it is permitted to do under the Electricity Act, and IESO is responsible to collect such fees and charges and pay them over to OPA. The charges in question expressly include those relating to procurement costs, which are deemed to have OEB approval: s 25.20 Electricity Act. As is the case with IESA, OPA may not set, change or eliminate charges without the approval of the OEB.

95 S 25.30 Electricity Act. The Minister of Energy may issue directives that OPA must take into account when developing the Integrated Power System Plan, including goals relating to the production of electricity from particular combinations of energy sources and generation technologies; increases in generation capacity from alternative energy sources, renewable energy sources or other energy sources; the phasing-out of coal-fired generation facilities; and the development and implementation of conservation measures, programs and targets on a system-wide basis or in particular service areas.

96 S 25.29.

This planning methodology is similar to the South African IRP process, but is more extensive, in that it provides not only for generation expansion but also investigates transmission expansion and the implementation of electricity conservation measures. It is submitted that the South African IRP planning process could only benefit from such an approach as increased generation capacity without appropriate transmission (ie conveyance) capacity serves no purpose.[97] It is also noteworthy that while IESO itself is not responsible for the IRP planning process, it is in the hands of a specialist agency and is not left generally to the devices to the Minister of Energy or his Department,[98] as in the South African context: section 4(1) of the ISMO Bill should thus be amended so that ISMO is not merely responsible for the modelling function but that it has full responsibility for Integrated Resource Planning as such.

4 4 2 System Expansions

The OPA must develop appropriate procurement processes for managing electricity supply, capacity and demand in accordance with its approved Integrated Power System Plans. These processes are subject to OEB approval, after which the OPA is free to enter into procurement contracts provided that it follows those processes as approved. No amendment can be made to the processes without OEB approval.[99] There is an obligation on the parties to a procurement contract to ensure that the contract provides a mechanism to resolve any disputes between them with respect to the contract.[100]

With reference to South Africa, it is clear that the allocation of new build opportunities (as identified in IRP planning) needs to be handled by an independent, neutral entity so as to ensure a level playing field. ISMO, under guidance of the Minister of Energy, should accordingly develop clear and transparent criteria for the allocation of new build opportunities to either Eskom or IPPs[101] and it has been suggested that a more robust approach such as the use of competitive bidding between Eskom and IPPs may be preferable.[102] It is quite apparent that Eskom is unable to undertake all investment needed for required new generation capacity – in fact, it has failed to add sufficient new capacity to adequately meet demand, as could be seen from the blackouts and load-shedding that faced the country in 2008, leading to the declaration of a national

97 See EIUG *op cit* 6.
98 Eberhard & Kapika argue that Energy Ministries rarely have adequate capacity to undertake detailed generation expansion planning and "Master Plans" either suffer from being too general or quickly become out of date. *Cf* McDaid *op cit*.
99 S 25.31.
100 S 25.32(3).
101 Eberhard & Kapika *op cit*.
102 Pickering *op cit* 4.

electricity emergency[103] – and also that it cannot necessarily deliver it more cheaply.[104]

IRP planning should furthermore be linked to timely initiation of appropriate generation procurement processes, preferably in the form of international competitive bidding processes.[105] Building long-term capacity for this within ISMO would be beneficial, and it would mean that generation procurement can be more easily linked to planning. Eberhard and Kapika submit that "procurement" in the Bill refers simply to buying and that the actual sourcing of power still vests in the Minister of Energy in terms of the ERA.[106] Although the ISMO Bill is congruent in this regard with the ERNGC promulgated in May 2011,[107] regulations can easily be substituted and it would be preferable that this be clarified in the ISMO Bill itself. The ERA would need to be amended accordingly.

4 4 3 System and Market Operator

The IESO may make rules governing the IESO-controlled grid; establishing and governing markets related to electricity and ancillary services; and establishing and enforcing standards and criteria relating to the reliability of electricity service or the IESO-controlled grid, including standards and criteria relating to electricity supply generated from sources connected to a distribution system that alone or in aggregate could impact the reliability of electricity service or the IESO-controlled grid.[108] The market rules may include provisions *inter alia* governing the conveying of electricity into, through or out of the IESO-controlled grid and the provision of ancillary services; authorising and governing the

103 See Eskom Briefing to Parliamentary Portfolio Committee on Minerals & Energy on "Current Electricity Crisis and Pricing" (2008-05-28). The Cape Town Regional Chamber of Commerce estimated a loss of R5.6 billion incurred by business in direct and consequential costs in respect of the Cape Town power disruption incidents that took place between Nov 2005 and Feb 2006, which resulted in the shut-down of Koeberg and subsequent load shedding due to insufficient reserve capacity: see *Energy Master Plan* 27 and Eskom Briefing to Parliamentary Portfolio Committee on Minerals & Energy on "Eskom's Response to Recent Cape Supply Interruption Incidents and Eskom's Build Programme" (2006-03-04). More recently, Eskom has been load shedding its larger customers due to the cold spell which saw eg peak demand forecasted at 36,258 MW with an available supply of 36,580 on 2012-06-16: see http://www.iol.co.za/business/business-news/eskom-cuts-power-demand-due-to-cold-spell-1.1343585 (accessed 2012-07-23), leaving a reserve margin of less than 1% – the *2007 Energy Master Plan* 31 set a target of 19% consistent with the 1 day in 10 year standard used in the US and many other industrialised countries.

104 See Eberhard & Kapika *op cit*.

105 *Ibid*.

106 *Ibid*. Pickering *op cit* 4 points out that "despite government's explicit 2001 policy to promote private investment into the sector a full six years have passed [since commencement of the Electricity Regulation Act] without the Minister making a single determination to allow private investment in generation capacity."

107 See 2 above.

108 S 32(1).

giving of directions by the IESO including, for the purpose of maintaining the reliability of electricity service or the IESO-controlled grid, such as to synchronise, desynchronise, increase, decrease or maintain electrical output, to take such other action as may be specified in the direction; authorising and governing the making of orders by the IESO, including orders imposing financial penalties on market participants, authorising a person to participate in the IESO-administered markets or to cause or permit electricity to be conveyed into, through or out of the IESO-controlled grid, or terminating, suspending or restricting a person's rights to participate in the IESO-administered markets or to cause or permit electricity to be conveyed into, through or out of the IESO-controlled grid.[109] The IESO may only make a market rule after first giving to the OEB an assessment of the impact of the rule on the interests of consumers with respect to prices and the reliability and quality of electricity service[110] and the OEB may also revoke any proposed amendment of a rule and refer it back to the IESO for its consideration.[111]

S 4(3) of the ISMO Bill is headed "Market Operator" but is mostly concerned with the contracting of new power. The making of any market rules would fall within the domain of NERSA. Buying of electricity generated is not a function of IESO, as Ontario has an open market system. As noted previously, OPA can contract for the procurement of new capacity – but there is no general rule that all new capacity has to be sold to OPA as in the single buyer South African model. Given that South Africa has no entity equivalent to OPA the ISMO Bill has rightly allocated the task to ISMO, who will sign long-term PPAs with IPPs. However, ISMO should not have the exclusive single-buyer function – there should be scope for IPPs to contract directly with eligible customers such as mines, industries and large municipalities, including non-exclusive cross-border trading.[112] This would give effect to Government's oft-repeated *dicta* in favour of introducing competition into the electricity supply industry in South Africa while at the same time providing IPPs with appropriate offtake guarantees to make their projects bankable.

4 5 Ownership of the Transmission Assets

As noted in 4 1 above, subsequent to the unbundling of Ontario Hydro the ownership and maintenance of the Ontario grid vests in Hydro One while OPG is responsible for generation. IESO therefore operates the grid but does not own it. Importantly, however, the grid is also not owned by an electricity generator but rather by a neutral entity.

109 S 32(2).
110 S 32(6).
111 S 33(3).
112 See Kruger *op cit*; Pickering *op cit*; EIUG *op cit*; Eberhard & Kapika *op cit*.

It was argued in section 3 above that it is vital that ownership of the transmission assets vests in ISMO, both for purposes of strengthening its balance sheet and to level the playing field between Eskom and the IPPs.[113] The position of the system operator would furthermore be difficult if the owner and maintainer of the national transmission grid is also an electricity generator[114] and it has therefore been suggested that neutrality requires that if the system does not belong to ISMO, it at least be sold off to some other independent, neutral entity, which could also present Eskom with a much needed capital injection.[115] With regards to renewable generation, Pickering argues that the location of primary renewable energy sources is often remote from existing infrastructure, which necessitates major investments in grid extension and strengthening. This, he cautions, would lead to a clear conflict of interest in Eskom's hands, as it is an active player in the generation sector and would therefore be strongly incentivised to favour its own projects at the cost of private projects.[116]

5 Conclusion

The ISMO Bill is an important step on the road to the introduction of competition in the South African electricity industry. It is urgent that the Bill be finalised so that industry players will have the necessary confidence to not only pledge, but actually invest the funds needed to get the Renewable Energy Procurement Programme underway. However, in the process of doing so, there exist serious areas of concern that need to be addressed in order to ensure a level playing field for the participants and also in the interests of securing the electricity capacity needed for medium and long term sustained growth and the development of a strong economy.

113 Also see EIUG *op cit* 7.
114 Kruger *op cit*.
115 See Free Market Institute *op cit*.
116 *Op cit* 7. Eskom has created the "Eskom Renewables Business" expressly for purposes of competing in the renewable energy sector: see Davenport "Eskom sets up Renewable Energy Unit" *Engineering News* (2011-03-15) available at http://www.engineeringnews.co.za/article/eskom-sets-up-renewable-energy-unit-2011-03-15 (accessed 2012-07-23).

The labour rights of educators in South Africa and Germany and quality education: An exploratory comparison

Johan Beckmann
BA THED BEd MEd DEd
Professor, Department of Education Management and Policy Studies, University of Pretoria

Hans-Peter Füssel
Ass Jur Dr Jur
Professor of Educational Governance, Humboldt University, Berlin and German Institute for International Educational Research Centre for Research on Educational Governance

OPSOMMING

Die Arbeidsregte van Opvoeders in Suid-Afrika en Duitsland en Kwaliteit-onderwys: 'n Verkennende Vergelyking

Die aandrang op toegang tot onderwys vir almal soos vergestalt in die *Education for All*-beweging van die 1990s het deur die jare verander in 'n aandrang op toegang vir almal tot kwaliteit basiese onderwys soos beoog in die *Dakar Framework for Action Education for All: Meeting Our Collective Commitments*. Wêreldwyd was daar al etlike pogings om kwaliteit-onderwys te definieer en dit word telkens in verband gebring met onder meer die kwaliteit van onderwysers en die gehalte van hulle opleiding, hoewel geen reglynige verhouding aangedui kan word nie. Die *UNESCO/ ILO Recommendation concerning the Status of Teachers* van 1966 gaan van die standpunt uit dat opvoeders toegang tot bepaalde regte, insluitende arbeidsregte, moet hê om hulle werk om kwaliteit-onderwys te voorsien behoorlik te kan doen. In Suid-Afrika het opvoeders volle eenvormige toegang tot die volledige reeks arbeidsregte in die Grondwet en ander wetgewing terwyl daar in Duitsland twee stelsels arbeidsregte vir opvoeders geld. Baie navorsing is nog nodig oor die verhouding tussen opvoeders se arbeidsregte (en die gebruik en misbruik daarvan) en die kwaliteit van onderwys en ons voer aan dat beperkings op fundamentele arbeidsregte en volle en onbeperkte toegang tot arbeidsregte sonder om enige ander faktore rakende kwaliteit-onderwys in ag te neem beide problematies is.

1 Introduction

Education for All (EFA) has been the focus of international education agencies for the past two decades. It involved providing education to everyone of appropriate age by 2015, thereby satisfying everyone's right to access to education. The challenge has been to simultaneously maintain and enhance the quality of education in schools across the globe. Howie[1] says that, after decades of deprivation in many contexts,

1 Howie expert lecture presented at the University of Pretoria on 2011-08-04.

it is important to not only discuss and reflect on how achievable quality education for all is, but also, what the cost is of not doing so. The concern about the quality of education has overtaken concerns about access to education and can be accompanied by questions about the status, rights, responsibilities and behaviour of teachers.

EFA was launched at the World Conference on Education held in Jomtien in Thailand in 1990. It was followed by a Mid-Decade Conference held in Amman in Jordan in 1996 and the World Education Forum held in Dakar, Senegal in 2000 where the *Dakar Framework for Action Education for All: Meeting Our Collective Commitments* was adopted. This document committed governments to achieving quality basic education for all by 2015 and represents both an up scaling of the Jomtien document (the word "quality" was added) and a down scaling (it is now "basic" education for all).[2]

Studies to measure the achievement of the goals of EFA and the *Framework* abound at both the international and national levels. PISA (the Programme for International Student Assessment), PIRLS (the Progress in International Reading Literacy Study), TIMMS (the Trends in International Mathematics and Science Study), SACMEQ (The Southern and Eastern Africa Consortium for Monitoring Educational Quality) and ANA (Annual National Assessment) are examples that come to mind.

PISA, TIMMS and PIRLs are international studies. South Africa does not participate in PISA but Germany does. SACMEQ is a regional study in Eastern and Southern Africa. ANA is a South Africa only national assessment of Grade 3 and Grade 6 learners' performance in language, literacy, numeracy and mathematics. The 2011 ANA study found the following national average performance levels: in Grade 3: literacy 35% and numeracy 28%; in grade 6: languages 28% and mathematics 30%.[3] South Africa has not done particularly well in any of the other studies and Germany does not top any of them either – the first results of PISA published in Germany in 2001[4] shocked the public and since then all new PISA results have been subjects of vibrant public debates in Germany.

In light of the above it is not surprising that there is a wide-spread effort to ensure the provision of quality education to all children.[5] Teachers are very important actors in regard to the quality and results of school education and one has to observe their role very carefully.

Because many role players impact the quality of education, one should guard against assuming a very direct causal relationship between the work of educators and the performance of learners. However, one may

2 Education for All http://www.unesco.org/education/efa/ed_for_all/ (accessed 2011-07-29).

3 In some instances teachers were not able to complete the papers intended for the students successfully.

4 Baumert *et al PISA 2000 Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (2001).

5 Fredriksson *Quality Education: The Key Role of Teachers* (2004).

identify conditions that are more or less important beneficial for the performance of teachers in schools. These conditions include the following:

i A common understanding and expectations of teachers' professionalism

The basis for any description of the role of teachers is a common understanding of their professionalism and of the expectations of, and the necessary conditions for their work as educators of the youth. Professionalism involves the behaviour, values and convictions shared by teachers.

ii Teacher education

The understanding of teachers' professionalism must also have an impact on teacher education. Of vital importance regarding teacher education are issues such as how one finds and attracts the most gifted future teachers, how the selection processes at universities must be organised to achieve the aim of finding the best human capital for the profession and how the content of teacher training as well as the practical component of university preparation for the teaching profession reflects an understanding of teachers' professionalism.

iii Understanding of teachers' status

The professionalism of teachers must be supported by the definition of teachers' status ie the (legal) rules describing the concrete rights and duties of teachers particularly in regard to their employment conditions. When looking at the concrete obligations of teachers one may ask if teachers are primarily "ordinary" employees with all the rights and duties laid down in normal labour law or whether there should be any restrictions, enhancements or embellishments of their labour rights because of their particular role and function as educators of the youth in a public school system.

These three ideas also suggest a certain sequential developmental dimension regarding teachers' professionalism:[6] An awareness of their professionalism, efforts to provide suitable teacher education (including further education and professional development) and the creation of conditions of service to support teachers to perform their primary task namely the education of learners optimally. The notion of a developmental dimension seems to be supported by the sequencing and structuring of the presentation of the rights of teachers in the *UNESCO/*

6 This notion was suggested to one of the authors by Dr Muavia Gallie, education consultant and one of the drafters of the framework for the continuing professional development of teachers in South Africa.

ILO Recommendation on the Status of Teachers[7] (the *Recommendation*) which was the first multi-national reflection on and articulation of the rights and status of teachers and which remains an important entry point into the debate on these two issues.

Clause 5 of the *Recommendation* provides that "[w]orking conditions for teachers should be such as will best promote effective learning and enable teachers to concentrate on their profession" while clause 79 provides that the

> participation of teachers in social and public life should be encouraged in the interests of the teacher´s personal development, of the education service and of society as a whole.

These clauses would seem to suggest that the drafters of the *Recommendation* linked teacher performance and their labour rights.

In the foregoing paragraphs we referred to increasing international insistence on the provision of quality education to all. We pointed out that teachers are assumed to play a vital role in the provision of education and that they seem to require access to certain labour rights in order to perform their task satisfactorily. The question then arose as to whether the nature of educators' work is such that it could justify measures to curtail or even withdraw some or all of their labour rights should they not contribute adequately to the provision of quality education or should they abuse their labour rights.

The above are the questions that we will examine in this article. In the paragraphs below we will provide an international perspective on some of the issues by referring to a number of provisions of the *Recommendation*, followed by an examination first of the situation as it obtains in South Africa and second of the *status quo* in Germany. We decided to compare South Africa and Germany for a number of reasons, among others the facts that:

(i) They are both constitutional democracies.

(ii) Teachers are employed by provinces or Länder.

(iii) Both countries have recently had to integrate disparate systems, and

(iv) Their education systems are performing differently in terms of international assessments.

In conclusion we will present tentative answers to our questions.

7 Adopted by the Special Intergovernmental Conference on the Status of Teachers, Paris, 19661005. See below in the section on international perspectives.

2 International Perspective

Education International[8] points out that the *Recommendation* is the only international standard applicable:

> [t]o all teachers in both public and private schools up to the completion of the secondary stage of education whether nursery, kindergarten, primary, intermediate or secondary, including those providing technical, vocational or art education.

The *Recommendation* consists of 146 clauses that set standards particularly in the following fields: Preparation for the profession, further education for teachers, employment and career, entry into the teaching profession, advancement and promotion, security of tenure, disciplinary procedures related to breaches of professional conduct, the rights and responsibilities of teachers, conditions for effective teaching and learning, teachers' salaries, and teacher shortages.

The *Recommendation* does not necessarily imply that teachers move hierarchically from the recognition of one category of rights to the next until they enjoy a complete set of working conditions and a status that "will best promote effective learning and enable teachers to concentrate on their professional tasks".[9] Although a developmental dimension is not spelt out explicitly, its existence whether categorical or implied is worth considering. The notion that teachers may have to earn the rights implied in the *Recommendation* sequentially may fly in the face of entitlement in terms of fundamental rights paradigms but one cannot escape the possibility that teachers might gain access to rights which they are not ready to exercise. In South Africa all teachers (including those that had been disadvantaged before) received the full array of labour rights referred to in the *Recommendation* in a single package in the wake of the 1994 transformation: Labour rights, rights to suitable professional preparation, the right to be heard, the right to professional development. The question arises as to whether such allocation has contributed to the delivery of quality education or whether the absence of some links in a chain has not negatively affected their service delivery.

It bears repeating that the link between quality education and the labour rights and status of teachers is not to be viewed naively and in an unsophisticated manner. Even though the *Recommendation* puts forward compelling arguments in favour of more rights for the majority of teachers in the world, it is still largely unimplemented in many corners of the globe and it would be wrong to suggest that all teachers across the world have access to all the rights, but the nexus between teachers' rights and educational quality seems firm and questions about the feasibility and viability of unbridled rights that may impair service delivery seem apposite.

---

8 Brochure on the 1966 *UNESCO/ILO Recommendation on the Status of Teachers* (2006).

9 *Recommendation* Cl 8.

In this contribution we will focus on teachers' labour rights. We will begin our discussion by examining the position in South Africa.

3 South Africa

The situation is South Africa is at the same time unique and new. The majority of teachers were deprived of many rights under the previous regime and all teachers instantaneously got access to all the fundamental rights entrenched in the Bill of Rights in chapter 2 of the Constitution, including the labour rights. Section 23 of the Constitution provides that:

> (1) Everyone has the right to fair labour practices.
>
> (2) Every worker has the right-
>
> (a) to form and join a trade union;
>
> (b) to participate in the activities and programmes of a trade union; and
>
> (c) to strike.

It should be noted that the Constitution refers to "workers" and not "employees". Teachers are therefore "workers" before the law and the decision to be classified as such was a conscious one and also displayed solidarity with the workers with whom the biggest South African teachers' union (the South African Democratic Teachers' Union (SADTU)) is in alliance.[10] Not all teachers' organisations are comfortable with the views and strategies of SADTU, which is often described as a hard line union, but in order to represent their members in negotiations and consultations they have no choice but to register as unions and take part in bargaining and negotiations even if they are more closely aligned to what might be called professional teachers' associations.

SADTU is a member of the Confederation of South African Trade Unions (COSATU) which is the largest union grouping in the country. Furthermore, COSATU is a member of the tripartite alliance which has united to form one group at the ballot box and it is, therefore, with the South African Communist Party an alliance which rules the country with the African National Congress (ANC) as ruling party. This situation where a trade union is part of government is unique and contradicts in many ways the notion that trade unions should primarily guard and promote their members' rights and interests.

The fact that teachers are at the same time employers and rulers is not without problems. For one thing, the governing party seems very hesitant to act against teachers who do not perform their duties as expected for fear of reprisals at the ballot box. One result is that protracted strikes (protected and unprotected) are the order of the day

10 This applies to teachers in the public and independent sectors alike. Unlike Germany (see below) South African law only recognises one group of teachers in the public sector.

and that the interests of children often play second fiddle to those of the teachers.

There is compelling anecdotal and other evidence that SADTU in particular is playing a disruptive role in schools predominantly serving poor communities. Motsohi refers to the debilitating and corrosive influence of SADTU.[11] Among the accusations levelled at SADTU are the following:[12]

(i) They place principals under enormous pressure and intimidation when they try to enforce normal learner and educator discipline.

(ii) They purport to be able to forbid principals to inspect teachers' work planning and to enter their classes.

(iii) It does not honour agreements to which it was a party.

(iv) They exert such pressure on principals that oppose them that such principals sometimes resign from education.

(v) They divide schools along party-political and union alliance lines and non-ANC and non-SADTU members are victimised. This includes fabricating rumours and charges of nepotism, financial and other mismanagement, racism and corruption.

(vi) They organise protests and force teachers and learners to participate in union activities which are sometime unauthorised and infringe on teaching time.

(vii) They close ranks when members are accused and found guilty of misconduct so that neither the principal nor the education authorities can or want to act against them.

If all of these allegations are true, it is clear that teachers' labour rights, if abused, may have an adverse effect on the quality of education and also on the management of educational institutions.[13] The situation is compounded by the fact that teachers chose not to get education declared as an essential service, which would have curtailed their labour rights to some degree. Although it is possible to limit teachers' labour rights in light of, and in accordance with section 36 of the Constitution and other legal principles, it would take exceptionally courageous political leadership to do so.[14]

---

11 "The corrosive influence of unions on SA schools" at http://education.ukzn.ac.za/News/11-08-22/The_corrosive_influence_of_unions_on_SA_schools (accessed 2011-09-01). See also "Lummel Sadou swaai die rottang in onderwys" at http://152.111.1.87/argief/berigte/rapport/ 2011/07/26/R2/III/24julonderwys.html (accessed 2011-09-15.)

12 Some would argue that SADTU is allowed to do certain things not because it is so strong but because the leadership in schools is poor.

13 At the moment there is a lack of large scale rigorous research to investigate the allegations of abuse of teachers' labour rights. However, allegations and anecdotes are persistent.

14 In a keynote address at the conference of the Distance Education and Teachers' Training (DETA) conference held in Maputo on 2011-08-2–5 the Vice-Chancellor of the University of the Free State and well-known

In South Africa one could argue that initial teacher education in South Africa is not in place yet as the *Policy on the Minimum Requirements for Teacher Education Qualifications*[15] has just been published. It provides "a basis for the construction of core curricula for initial teacher education, as well as for Continuing Professional Development (CPD) programmes[16] that accredited institutions must use in order to develop their programmes leading to teacher education qualifications. The tone of the *Policy* suggests that the teacher education contemplated must still happen and that teachers' rights to such education exist but have not been given full expression as yet. It would seem anomalous then to award teachers the right to continuing professional development, which is by nature focused on honing already-acquired competencies, skills, knowledge and values as a teacher, at the time of the release of this future-oriented statement on the rights of educators. The *Policy* also intimates that the right to initial teacher education precedes the right to continuing education and that continuing education in the form of re-skilling should normally not be used a way to rectify dysfunctional or ineffective training. Yet the *National Policy Framework for Teacher Education and Development* (NPFTED) of 26 April 2007 makes provision for a national system of Continuing Professional Teacher Development (CPTD). The issue was again debated comprehensively at the national Teacher Development Summit held from 29 June to 2 July 2009.

The foregoing paragraphs suggest therefore that there may be grounds for limiting teachers' rights including their labour rights on the basis of the current stage of development of teaching as a profession. The argument departs from the *Recommendation*'s separation of four types of teachers' rights, namely those regarding initial teacher education, professional registration, professional development and participation in negotiations (through unions). The above paragraphs propose that one needs to examine the labour rights of teachers more closely in light of teaching's incomplete compliance with the characteristics of a profession as set out in the *Recommendation* and in light of the poor performance of the South African education system.

The graph in Figure 1 depicts the minimum standards for primary teaching and number of teachers who met them in Sub-Saharan Africa in 2006.[17] International Standard Classification of Education (ISCED) 2 represents 8-9 years of schooling; ISCED 3 represents 12-13 years of

educational and social commentator, ascribed the very poor performance of South African education to the absence of such leadership in education and in politics in the country.

15 Department of Higher Education and Training *Policy on the Minimum Requirements for Teacher Education Qualifications* GN 583 GG 34467 of 2011-07-15.

16 *Idem* 8.

17 Global Campaign for Education "Teachers For All: what governments and donors should do. Policy briefing of the Global Campaign for Education" 30-33 at http:// www.vsointernational.org/Images/GCE_Teachers_For_All_tcm 76-22710.pdf (accessed 2011-07-12).

schooling (upper secondary); ISCED 4 represents bridging education between secondary and post-secondary education; and ISCED 5 represents 2 – 3 years post-secondary education.

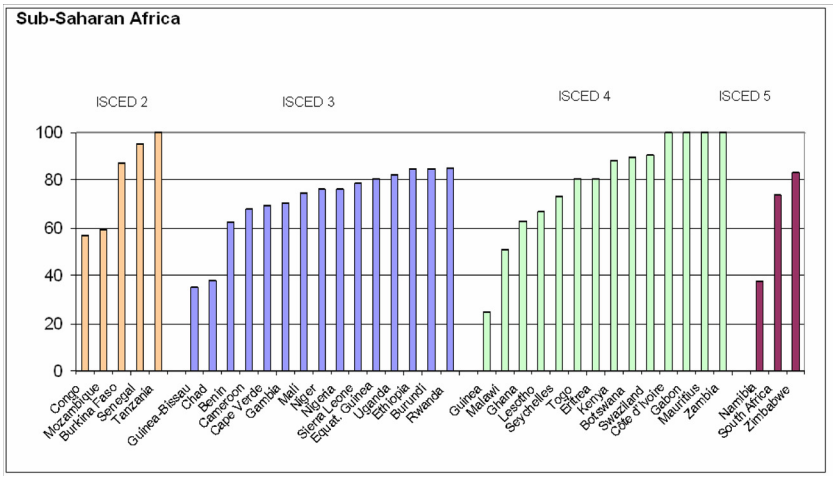


Figure 1: Minimum standards for primary teaching and number of teachers who met them in Sub-Saharan Africa in 2006 (Global Campaign for Education, 2006)

South Africa has the second highest number of teachers meeting ISCED in Sub-Saharan Africa and expectations that the national education system will perform well in comparison with other countries in the region are therefore justified.

We will now proceed to a discussion of South African teachers' labour rights.

4 Labour rights of South African teachers

4 1 Labour Relations Act

The Labour Relations Act[18] (LRA) was promulgated among others to give effect to the public international law obligations of the Republic relating to labour relations,[19] to give effect to section 23 of the Constitution; to promote and facilitate collective bargaining at the workplace and at sectoral level; to promote employee participation in decision-making

18 66 of 1995.

19 This seems to bring teachers' labour law rights within the ambit of the *Recommendation* although it does not enjoy the status of international law acceded to by member countries.

through the establishment of workplace forums; and to provide simple procedures for the resolution of labour disputes through statutory conciliation, mediation and arbitration (for which purpose the Commission for Conciliation, Mediation and Arbitration (CCMA) is established), and through independent alternative dispute resolution services accredited for that purpose. The LRA specifically protects employees against unfair dismissals.

Teachers therefore have access to collective bargaining, to participation in decision making and to conciliation, mediation and arbitration for the resolution of labour disputes. In terms of Item 3(2) of Schedule 1 to the LRA the Education Labour Relations Council (ELRC) (established by section 6(1) of the Education Labour Relations Act)[20] is deemed to be a bargaining council established in terms of section 37(3)(b) of the LRA. The agreements adopted by this council are binding on employers and employees.[21]

4 2 Employment of Educators Act

The Employment of Educators Act[22] (EEA) provides for the employment of educators by the State, for the regulation of their conditions of service, for discipline, for retirement and for the discharge of educators.[23] The EEA defines who educators and employers are[24] and it thus very clear to whom the rights contained in this law accrue.

A number of regulations have been published under the EEA including Conditions of Service of Educators, Determination of Salary Adjustments for Educators, Personnel Administration Measures (PAM), Regulations to Provide for Interim Measures according to which Rationalisation in Education in terms of Resolution 3 of 1996 and Other Related Agreements of the Education Labour Relations Council can be completed, the Role of Managers Prior to Strike Action and Regulations regarding the Staffing of Rationalised Educational Institutions. All of these create enforceable rights and obligations of teachers.

Section 2 of the EEA provides that the EEA applies to the employment of educators at (a) public schools; (b) further education and training institutions; (c) departmental offices; and (d) adult basic education centres. Section 3 spells out who the employers of specific categories of employees are for all purposes of employment as well as for the purposes

20 This Act was replaced by the Educators Employment Act, 1994 which was superseded by the Employment of Educators Act 76 of 1998.
21 S 23 of the LRA.
22 76 of 1998.
23 Teachers enjoy such comprehensive protection against amongst others unfair labour practices through a wide array of legislation that disciplinary action against them presents a serious challenge to managers. It is not impossible that some teachers may indeed consider themselves beyond the reach of the law. Also see Motsohi *op cit* above.
24 S 1 EEA.

of determining the salaries and other conditions of service of educators and for the purposes of creating posts.

Chapter 3 of the EEA deals with the appointments, promotions and transfers of educators and section 6 provides for the powers of employers. Section 6(1)(b) provides that the appointment, transfer or promotion of an educator in the service of a provincial department shall be made by the provincial Head of Department (HoD). Section 3(a) provides that such appointment, promotion or transfer may only be made on the recommendation of the governing body (SGB) of the public school. Section 3(b) contains instructions with which governing bodies should comply when considering applications with a view to preparing recommendations for the employer. Section 6(3)(c) provides that the governing body must submit, in order of preference to the HoD, a list of – (i) at least three names of recommended candidates; or (ii) fewer than three candidates in consultation with the HoD.

Sections 6(3)(d) to (g) outline the responsibilities of the HoD in considering and dealing with the recommendations contemplated in sections 6(3)(a) to (c). Section 3(f) provides that, despite the order of preference in paragraph (c) and subject to paragraph (d), the HoD may appoint any suitable candidate on the list. Section 3(g)(iii) provides that, when an HoD declines a governing body recommendation he or she must, despite section 6(3)(a), appoint a suitable candidate temporarily or re-advertise the post.

The power given to an HoD to appoint someone despite the SGB's recommendations may constitute a violation of the rights of both SGBs and teachers. The courts do not seem to view the power as a carte blanche given to HoDs. In *Settlers Agricultural High School v Head of Department: Department of Education, Limpopo Province*[25] the SGB of the school had recommended the appointment of a candidate as principal. The HoD rejected the recommendation and appointed another candidate in the post. The school and its SGB successfully obtained the overturning of that decision by the court and the respondents appealed for leave to appeal the overturning.

Without making a finding Bertelsmann J[26] stated the following:

> It might possibly be argued … that the provisions of section 7(1)(b) indicate that a candidate from a previously disadvantaged community ought to be preferred in cases where the evaluation of such candidate and a competitor from a previously privileged group leads to a comparative parity in the assessment of their suitability for the post. But where the difference in the respective suitability for the post is, in the opinion of an interviewing committee which honestly applies the agreed procedure, as substantial as is the case here, neither the Constitution nor the statute or the equity plan [in terms of the Employment Equity Act 55 of 1998] demand the preferring of

25 [2002] JOL 10167 (T).
26 22-26.

> the candidate who belongs to a group which was previously discriminated against.

But the Constitution also entrenches the right to proper education and provides specific protection for children. Section 28(2) of the Constitution reads as follows:

> 'A child's best interests are of paramount importance in every matter concerning the child.'
> As important as the rights of educators, and in particular those belonging to previously disadvantaged communities are, the paramountcy of children's rights and interests must not be overlooked. I am of the view that the first respondent [the HoD of the particular provincial department of education] would not be entitled to substitute his own choice for that of the interviewing committee and the school governing body, but would have to refer the matter back to the interviewing committee and the school governing body with the instruction to apply the law properly.[27] Section 6(3)(c) of the [Employment of Educators] Act appears to be clear in this regard.

Bertelsmann J consequently refused the leave to appeal with costs.[28] It seems then that in this instance the court upheld the rights of the teacher concerned to just administrative action. In this regard Section 33(1) of the Constitution reads as follows: "(1) Everyone has the right to administrative action that is lawful, reasonable and procedurally fair." The action of the HoD in question was also not in line with section 6 of the EEA.

One other example will suffice to illustrate the protection that teachers enjoy against arbitrary limitations of their labour. In *Head of Western Cape Education Department v Governing Body of Point High School*[29] Hurt AJA commented that, if the HoD "considers that the governing body has performed its functions properly, the HoD must attribute substantial weight to the recommendations submitted to him".[30] In this case the HoD of the Western Cape Department of Education declined the recommendations of the SGB for the school regarding the appointments of a principal and deputy-principal respectively. Hurt AJA commented that the assumption that the scope of the discretion of the HoD had been broadened was obviously correct but noted that this did not excuse him from "having to furnish acceptable reasons for his decision".[31]

It is worth noting that the candidates recommended by the SGB and those appointed by the HoD were all white males, a group over-represented in the Western Cape at the respective post levels. The two candidates recommended by the SGB were both from the KwaZulu-Natal

27 This aspect of the EEA was later amended materially, as the wording of s 6(3)(f) clearly shows, "... despite the order of preference in paragraph (c) and subject to paragraph (d), the Head of Department may appoint any suitable candidate on the list".
28 26.
29 [2008] ZASCA 48.
30 12.
31 13.

province. When the SGB asked for reasons why their recommendation for deputy-principal had not been appointed, the Department responded as follows in a letter dated 4 December 2006:[32]

> As you are aware, there is an over-representation of males at post level three in the WCED. The appointment of any of the nominees would not have promoted or improved the [Employment Equity] targets of the WCED, therefore the appointment of Mr Swanepoel was approved.

The Department argued that if they appointed a person from the Western Cape, vacancies would arise which could be used for EE purposes.[33] The judge pointed out that:

> [the HoD] failed, signally, to perform the balancing exercise referrred to in *Bata Star* by weighing the (somewhat obscure) employment equity considerations which had occurred to him, against the disparity and suitability between the candidates recommended by the Governing Body and the candidates whom he decided to appoint.

He went on to point that employment "equity provisions should only prevail where there is approximate equality between the ability or ability of the two cases".[34] The judge consequently set aside the HoD's decisions:

> [o]n the broad ground of unreasonableness as contemplated in s 6(2)(h) [of the EEA]. In my view the HoD proceeded without a proper understanding of the discretion which he was called upon to exercise.[35]

Although we explore, in this article, the possibility that South African educators' full array of labour and other rights may not be justified and may be counterproductive in regard to quality education because of their abuse by educators, some of them do positively contribute to quality education and need to be applauded. The cases discussed above are cases in point where the application of the law probably ensured that the best possible candidates for specific posts were appointed.

In addition to a myriad of disputes brought before the labour dispute resolution mechanisms of mediation and arbitration, the CCMA, the Labour Court and the Labour Court of Appeal, there are many other examples of cases in which various teachers' rights were adjudicated, including:[36]

(i) *Mangela v MEC, Department of Education, EC*[37] (termination of employment of a temporary educator).

---

32 7.
33 9.
34 In this case there was clearly no approximate equality of ability as appeared from the evidence put forward by the SGB.
35 18.
36 Information obtainable on the website of the Federation of South African Schools (FEDSAS) (www.fedsas.org.za (accessed 2011-07-12)) whose summary of the litigation in question we have used.
37 [2006] ZAECHC 41.

(ii) *Nakin v MEC. Department of Education, Eastern Cape Province*[38] (transfer of principal).

(iii) *Phenithi v Minister of Education*[39] (dismissal of educator).

(iv) *Reddy v KZN Department of Education and Culture*[40] (unfair labour practice).

(v) *MEC for Education, Western Cape Province v Strauss*[41] (injury on duty).

(vi) *Larbi-Odam v Member of the Executive Council for Education (North-West Province)*[42] (intended termination of services of temporarily-employed expatriate teachers).

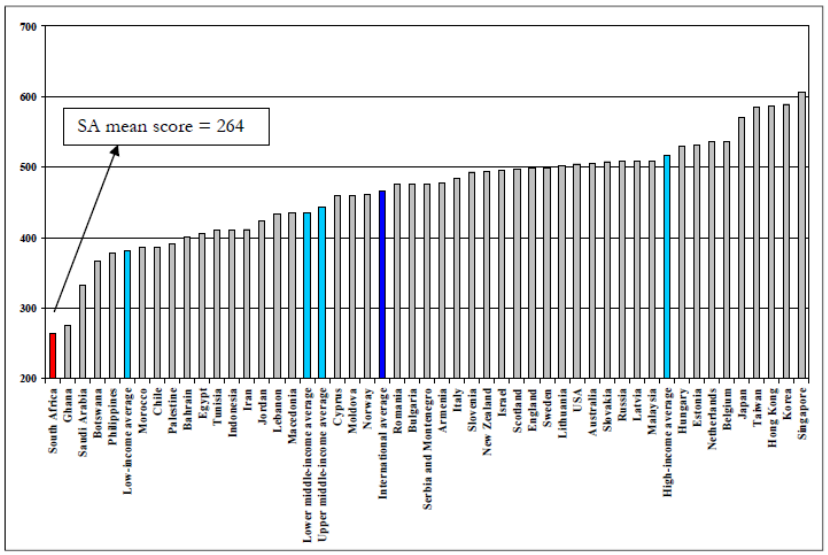
After discussing the rights and status of teachers in South Africa in the wake of the new political dispensation after 1994, we will now make some broad comments on the quality of education available to the country's children since 1994.

4 3 The Quality of Education in South Africa After 1994

Despite the significant number of rights which teachers have gained since 1994, no significant advances in education quality (especially for the poorer parts of the population concentrated in rural areas) since then can be indicated. Spaull[43] points out that, given the racial dimension of poverty, and that the poor are more likely to be black, it appears that, on average, black students receive an inferior quality of education to their white peers. He comments that the fact that this "is the reality 17 years on from apartheid is particularly disconcerting." Especially for black children in rural areas, 17 years of comprehensive teacher rights have not changed their plight.

Even more disconcerting is the representation by Taylor *et al*[44] of South Africa's mathematics performance in TIMMS 2002.[45] The graph in Figure 2 requires no elucidation.

38 [2008] ZAECHC 13.
39 [2003] ZACC 16.
40 [2003] ZALAC 7.
41 [2007] ZASCA 155.
42 [1997] ZACC 16.
43 Spaull *A Preliminary Analysis of SACMEQ III South Africa* (2011).
44 Taylor, Van der Berg, Reddy and Janse van Rensburg *How well do South African schools convert grade 8 achievement into matric outcomes?* (2011) 7.
45 6 African countries were included in the 2002 TIMMS survey.



Note: The TIMSS scores are scale average scores set to have an international mean of 500 and standard deviation of 100.

Figure 2: South Africa's mathematics performance in TIMMS 2002[46]

Taylor, Van der Berg, Reddy and Janse van Rensburg point out that in the 2000 and 2007 SACMEQ surveys of reading and mathematics, South African children performed below average in comparison with other Southern and East African countries.[47] Given that South Africa has more qualified teachers, lower pupil-to-teacher-ratios and better access to resources, one would expect that South African students would perform at the top of the regional distribution. Unfortunately this is not the case. In a league table of student performance,[48] South Africa ranks 10th out of the 15 SACMEQ countries for student reading performance and 8th out of 15 for student maths performance.[49] Spaull[50] points out a SACMEQ result which is hard to comprehend:

> [s]ince the teacher tests contained many of the same questions as the student tests, one would expect all teachers to score almost full marks on the teacher test. This was certainly not the case.

46 Taylor *et al* 7.
47 Spaull 24
48 23.
49 Spaull 23-24.
50 22

One wonders if South African teachers have fully achieved the first requirement of the status recommendation, namely that of having adequate professional qualifications.

Spaull[51] concludes:

> While the constitution promises equal access to education, it cannot promise an equal quality of education. Until such a time as the primary education system in South Africa is able to offer *a quality education to all students*,[52] not only the wealthy, the existing levels of educational inequality will remain.

Taylor[53] poses a question embedded in a statement that echoes Spaull's assessment of the quality of education:

> Inequity in the quality of education has proved a more enduring problem. For many poor children, who are predominantly located in the historically disadvantaged part of the school system, this *low quality of education* acts as a poverty trap by precluding them from achieving the level of educational outcomes necessary to be competitive in the labour market. An important question is the extent to which this low quality of education is attributable to poverty itself as opposed to other features of teaching and management that characterise these schools.
>
> The substantial increase in resources invested in the historically disadvantaged parts of the school system has unfortunately not produced a commensurate improvement in education quality. This is clearly evident in the test scores of South African students in numerous surveys [such as TIMMS, PIRLS and SACMEQ] of educational achievement that have been carried out in recent years These surveys have unequivocally shown that *the overall level of achievement amongst South African children is extremely low*.[54]

Taylor's question leads to other questions. Among these are: If the status and rights of teachers do not seem to be contributing to providing quality education to all the learners, should such rights and status be examined with a view to such adaptation as might facilitate the improvement of the quality of education. We will attempt a provisional answer in the conclusion of this contribution.

First we will now sketch the situation regarding teachers' rights and status in Germany.

51 26.
52 Our emphasis.
53 Taylor *Uncovering indicators of effective school management in South Africa using the National School Effectiveness Study* (2011) 3-4.
54 Our emphasis.

5 Germany

5 1 Introduction[55]

In Germany the responsibility for the education system is part of the authority of the sixteen regions or *Länder*.[56] The Federation or *Bund* only provides the framework for the legislation of the regions. Therefore the status of the teachers is ruled by regional law: Teacher education takes place in terms of regional law, the recruitment of teachers is the task of the regional Ministries of Education, the allocation of teachers is organised by these regional Ministries and their subunits and teachers are remunerated out of the budgets of the regions.

The federal framework for the action (authority) of the regions is primarily provided by the Federal Basic Law or *Grundgesetz*. On the one hand civil rights limit the actions of the regions and on the other hand some areas of legislative competence (authority) is the responsibility of the Federation on the basis of a common and nationwide equal ruling.[57] The legal status of teachers is regulated by federal as well as regional law.

5 2 Kinds of legal status of teachers

In Germany all teachers at public schools are civil servants, but there are two different kinds of legal status for teachers: as public officers (*Beamte*) and as employees (*Beschäftigte*). The legal status of the first group is defined by public law and that of the second group of teachers by civil- and labour law.

Which of these two different kinds of legal status is "right" and "adequate"? That question is first and foremost a constitutional one to be decided in terms of the rules of the *Grundgesetz*. Article 33 Paragraph 4 of the *Grundgesetz* reads as follows in this regard:

> The exercise of sovereign authority on a regular basis shall, as a rule, be entrusted to members of the public service who stand in a relationship of service and loyalty defined by public law.

This juridical question is therefore whether or not teachers exercise sovereign authority and whether or not they stand in a relationship of service and loyalty to the state as defined by public law.

---

55 Quotations from the German *Grundgesetz* are from official translations into English; all other translations by the authors.

56 See Art 70 Par 1 *Grundgesetz*: "The *Länder* shall have the right to legislate insofar as this Basic Law does not confer legislative power on the Federation".

57 See Art 72 Par 2 *Grundgesetz*: "The Federation shall have the right to legislate on matters … if and to the extent that the establishment of equivalent living conditions throughout the federal territory or the maintenance of legal or economic unity renders federal regulation necessary in the national interest".

If it can be said that teaching of the youth is of exceptional importance to society and that the state depends on a well-educated population and that the task of preparing the youth to be good citizens can only be fulfilled in a relationship of service and loyalty to the state and must be understood as a means of exercising sovereign authority. This would also be true of teachers in private schools who must then be regarded as public officers because they also educate citizens.

The difficulties associated with this broad interpretation of the exercising of sovereign authority have led to a reduction of the role of teachers, confining it to the ends of learners' school-careers during examinations. It is argued that examinations and their results are of such importance to the youth and their future life that only the state can be allowed to decide on a matter like this. If teachers are involved in decisions in these matters their acts must be qualified as exercising sovereign authority. Teachers are therefore regarded as public officers under the public law regime. As in most juridical debates this position is not acknowledged by all. The German Courts have not given a ruling on this dispute yet.

If one introduces European law and the jurisdiction of the European Court of Justice into the debate, it seems that the general understanding of exercising sovereign authority does not capture and describe the essence of teachers' work and profession adequately. In terms of European law only the work done by officials (bureaucrats) in state departments as well as the police and other state agencies falls within the ambit of exercising of sovereign authority. How far the broader European understanding of legal constructs may or could influence the German national interpretation of concepts is a legal question that has also not yet been resolved.

Because of the juridical indecision the question regarding the "right" and "adequate" status for teachers is decided by other criteria – financial ones. As has already been mentioned, both groups of teachers (the *Beamte* and the *Beschäftigte*) are employed and paid by the regional governments. They pay teachers' salaries and, as long as teachers are public servants, their contributions towards their retirement pensions are also paid by the regions. If teachers are employed under a civil/labour law dispensation, the pensions for the retired people will be paid by special entities, the social security institutions, and not by the regions. That means that teachers employed under labour law attract less government expenditure in the long run than the teachers who are public servants. So the open juridical question is decided by financial arguments.

There is a common understanding that regional governments are more concerned about money and finances than upholding constitutional aspects especially in cases where legal answers and interpretations are not very clear as in this case. When deciding on the legal status of teachers, governments of regions tend to be dictated to by financial matters as one can see in the regions of the eastern part of

Germany, the former German Democratic Republic, where most of the teachers do not have the status of public servants (except for headmasters and headmistresses).

As the legal debate has not been resolved, both forms of status are accepted – and often teachers with different legal status work together in the same school, some with the status of public officers (*Beamte*) and some with the status of employees (*Beschäftigte*) doing the same work. This confirms again that finances tend to have a stronger influence than the law on decisions regarding the status of teachers.

The fact remains that the majority of teachers are public officers (*Beamte*) with rights and duties laid down in public law but the Region of Berlin, for instance, is no longer employing new teachers as public officers. Consequently applicants move or migrate towards regions where they still enjoy the status as public officers (*Beamte*).

5 3 The Legal Status of Teachers as Public Officers (*Beamte*)[58]

5 3 1 Federal Constitutional Conditions

The general framework for the legal status of public officers in Germany – and this includes teachers - is described in Article 33 Paragraph 5 of the *Grundgesetz*:

> The law governing the public service shall be regulated and developed with due regard to the traditional principles of the professional public service.

This rule is understood and applied in three manners: legislative demands are formulated, the system of public officers (*Beamte*) as such is guaranteed by the *Grundgesetz* and this clause may also limit rights of public officers.

The *Grundgesetz* assigns the competence to regulate the general conditions of rights and duties of public officers to the Federation.[59] The Federal Act on the Status of Public Officers (*Beamtenstatutsgesetz*) was enacted in terms of this competence but this federal law is only the framework for the Acts on Public Officers of the Regions (*Landesbeamtengesetz*).

The School Acts of the regions elaborate more on teachers' duties. The legal structure underpinning the legal status of teachers can therefore be depicted as follows:

58 See also Füssel "Nichtvermögenswerte Rechte der Lehrkräfte" in *Schulrecht* (eds Avenarius & Füssel) (2010) 634*ff*.
59 Art 74 Par 1 No 27 *Grundgesetz*.

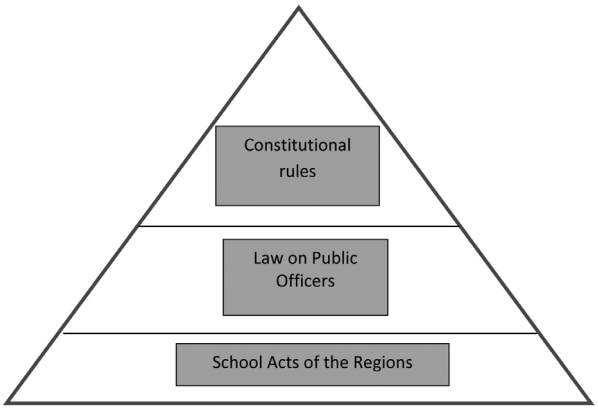


Figure 3: Pyramid of legal rules determining the status of German teachers as Public Officers (Beamte)

The constitutional clause on "the traditional principles of the professional public service" has been developed over time by the courts. In terms of the law a restriction of the fundamental rights of Public Officers is acceptable as long and insofar as these limitations are necessary and justifiable in the context of their special tasks and obligations in the service of the public. For example: the *Grundgesetz* guarantees the freedom of expression in Article 5 Paragraph 1 but using this general right may cause problems regarding the interests of the public demanding neutrality of public officers. So a restriction of the freedom of expression is accepted by referring to Article 33 Paragraph 5 of the *Grundgesetz*. The Federal Constitutional Court is enjoined to find a solution for conflicts like these by way of "practical concordance/ agreement" which means that, in each specific case, ways of accepting/ reconciling the different interests must be developed by looking for solutions that comply with both clauses as far as possible.

What a "practical concordance/agreement" may look like can be seen in the clauses of the Federal Act on the Status of Public Officers dealing with fundamental rights.

5 3 2 Federal Act on the Status of Public Officers[60]

In regard to the resolution of conflict between the freedom of expression and the duties of Public Officers by using the principle of practical concordance one may consult Paragraph 33 of the Federal Act on the Status of Public Officers which provides that:

60 Dated 2008-06-17/2009-02-05 (BGBl.2008 I S. 1010; BGBl. l 2009 I S. 160).

> Public servants must exercise care in political acting in a way of moderation and restraint which follows out of their position in direction of the common public and in respect of the duties of their functions

and

Paragraph 34 of the Federal Act on the Status of the Public Servants which requires: "The behaviour of public servants … [to] reflect the respect and the reliability which are necessary to fulfil the duties of their profession".

Both clauses must be understood as an attempt to constitute a general rule for dealing with the conflict between the fundamental right to freedom of expression as laid down in Article 5 Paragraph 1 of the *Grundgesetz* on one hand and the limitation set by the "traditional principles of the professional public service" (Article 33 Paragraph 5 of the *Grundgesetz*) on the other hand. This regulation applies to all public officers, to persons working in the ministries, to policemen and also to teachers.

The general phrasing of the above clauses suggests that they could also be used to solve other conflicts concerning fundamental rights, for instance disputes in regard to freedom of religion or freedom of assembly.

5 3 3 Acts on Public Officers of the Regions

As has already been indicated teachers are employed and remunerated by the regions. Therefore their rights and duties are determined by the law enacted by the regions, especially the Acts on Public Officers of the Regions. In terms of the framework of the Federal Law the regions are, however, bound regarding their legislation by the concrete rules of the Federal Law; so the regional Acts on Public Officers generally echo what has been said already in the Federal Law.

5 3 4 School Acts of the Regions

It is not only the general regulations for Public Officers that apply to teachers of that status but the School Acts of the Regions (*Landesschulgesetze*) also embody further regulations concerning the behaviour of teachers.

In terms of the regional School Acts schools have to comply with the principles of neutrality and tolerance. Teachers' conduct in schools must therefore be informed by neutrality and tolerance. In teaching and educating young persons on the principles of neutrality and tolerance teachers must act as models of these concepts, so that pupils can also learn these general aspects in their daily school life.

A rule like Paragraph 67 of the School Act of Berlin [61] which provides that:

> [t]eachers have, beside their right to voice their own opinion while teaching, also to ensure that other opinions, which are of importance to education and to the tasks of the school, are heard; any one-sided manipulation of the pupils is forbidden

is legally acceptable even if it restricts the personal fundamental rights of a teacher. The justification for that restriction can be found in the general task of the school namely to educate young persons.

This restriction of the action and conduct of teachers also seems to be appropriate in light of the principle of neutrality and tolerance[62] necessary for schooling. It is therefore not only the balancing of legal aspects that constrains teacher conduct and necessitates the acceptance of limitations of their fundamental rights but also relevant pedagogical aspects.

5 3 5 The Legal Status of Teachers as Employees (Beschäftigte)[63]

As has already been pointed out above the legal status of teachers in Germany may also be that of as employees; hence labour- and civil law applies to them. The concrete duties and rights of teachers in terms of such status follow their working contracts which are embedded in general Labour Law regulations and especially the Collective Labour Agreements for Civil Servants concluded between the regions as employers and the unions in the public sector.[64] There are general provisions on the duty of employees to act in loyalty to their employer and that rule includes the duty to accept the principle of neutrality and tolerance in schools.[65]

The School Acts of the Regions do not distinguish between the different types of teacher status; so the clauses quoted already are also applicable in the Regions.

5 4 Case Law

It is to be expected that the definition of concrete rights and duties of teachers has been developed by the courts. We refer to some of the important cases below.

61 Dated 2004-01-26/2011-07-13 (GVBl: 2004, S, 26 / GVBl: 2011, S,347).
62 See Federal Constitutional Court decision dated 1979-10-16 (BVerfGE 52, 223 (247)).
63 See also Füssel 732*ff*.
64 "Tarifvertrag für den öffentlichen Dienst" (2011-03-10).
65 See Federal Administrative Court decision dated 1981-01-19 (BVerwGE 81, 212).

5 4 1 Freedom of Assembly

The right to freedom of assembly (Article 8 of the *Grundgesetz*) includes the right to decide where and when an assembly should or may take place. An assembly might clash with the duties of teachers to stay on the school grounds during times scheduled for teaching. Here the courts have ruled that participating in assemblies is normally only possible in times outside the working time of teachers – in these cases the "practical concordance" implies that the right of assembly as such is accepted for teachers too but the right to decide at which time an assembly will or may take place is limited.[66]

5 4 2 Freedom of Association

The freedom of association in terms of Article 9 of the *Grundgesetz* includes in Paragraph 3 "[t]he right to form associations to safeguard and improve working and economic conditions [for] every individual and every occupation or profession".

The clause is interpreted in a manner that includes the rights to go on strike.[67] This suggests that teachers also have a right to go on strike. In terms of a strict interpretation of the clause "every occupation or profession" all teachers should be included. However, in light of the abovementioned constitutional clause that the law of public officers should be "regulated with due regard to the traditional principles of the professional public service" (Article 33 Paragraph 5 of the *Grundgesetz*) the courts have ruled and accepted that in general public officers do not have the right to go on strike,[68] and that includes teachers as far as they are public officers under the Law on Public Servants regime. However, for contract teachers (employees) the right to go on strike is not limited. One can therefore not categorically claim that, because of the general tasks of schooling, teacher strikes are generally prohibited.

5 4 3 Multiple Roles of Teachers

The above examples of the juridical interpretation of the status of teachers in Germany can also be construed to imply that the status and work of as public officers (*Beamte*) are less important than the educational mission / assignment of the teacher. On the contrary, being an employee (*Beschäftigte*) may signify that teachers' educational mission is of no consequence or special importance; their legal status is comparable with any worker in an office or a factory in terms of the law. One would have thought that the government would make a decision in favour of either *Beamte* or *Beschäftigte* on constitutional or legal grounds.

---

66 See Administrative High Court of Northrhine-Westfalia decision dated 1995-05-17 (SPE nF 395 No 2).

67 Federal Constitutional Court decision dated 1993-03-02 (BVerfGE 88, 103).

68 Federal Constitutional Court decision dated 1951-10-23 (BVerfGE 8, 1); Federal Constitutional Court decision dated 1997-03-30 (BVerfGE 44, 249); Federal Administrative Court decision dated 1980-12-03 (BVerwGE 73, 97)

However, governments seem to make decisions in this regard on fiscal reasons only: They accept the different kinds of status for teachers which are not system based and for which there are no legal reasons.

The results of Large Scale Assessments like PISA and TIMMS in the schools of the regions where this seemingly anomalous situation of different teacher statuses even in the same school exists are in general not worse or better than in other parts of the country. It would therefore seem that status does not impact the results of teaching and learning significantly.

A by-product of these different forms of status accorded to teachers in general is a degree of competition between the regions for the services of teachers. Different regions may not only offer a different status to teachers but the salaries (and the pensions) may also be different. This problem will become more serious in the future and may cause serious problems for the poorer regions in recruiting teachers. A shortage of teachers is likely to impact the results of teaching and learning negatively.

It appears that in Germany the "right" balance between the role of teachers as educators in state-run schools and their acting in the public interest on one hand and the status of teachers as either public officers (with limited rights) or as "ordinary" employees (with full rights) on the other hand is not yet clearly defined. Perhaps a kind of "double role" of teachers needs to be considered and accepted including that the roles should be equally important and that one should not be subject to the other. As far as the general task of schooling, namely education as performed is concerned, it appears that the balancing of interests in light of the principles of neutrality and tolerance can also be interpreted as a form of education, as education by example. Debates about the "right" legal status of teachers must also be seen in this light and be cognisant of the fulfilment of the school's tasks and aims.

It seems that a general solution of the conflicts contemplated in the *UNESCO/ILO Recommendation* is not attainable in Germany, perhaps because the importance of the various conflicts is beyond the reach of general regulation.

6 Conclusion

We set out to look for a link between the rights of teachers and the quality of education. In Germany teachers employed by the same regional government may work side by side in the same public school but under two inexplicably different employment regimes save for fiscal reasons[69] – one group having access to more rights (including the right to strike) than the other (who may even be paid less). We found that, in Germany,

69 In South Africa, with its understandable obsession with equality and the prevention of unfair discrimination, such a dispensation is unconscionable.

the existence of two employment regimes (even in the same public school) does not appear to have a significant impact on the quality of education and no differentiations of results could be linked to different rights regimes. The German approach seems to pursue a balancing of rights rather than the total or partial limitation or rescinding of rights.

South African educators have almost unconditional access to the fullest possible labour and other rights subject to restriction in terms of constitutional provisions. This recently-gained access to the full range of rights with an enormous increase in resources has not been accompanied by a demonstrated improvement in the quality of education to the country's children. Van der Berg[70] summarises the views of a number of commentators succinctly:

> Despite narrowing attainment differentials, unprecedented resource transfers to black schools and large inflows of black pupils to historically white schools, studies have shown that historically white and Indian schools still far outperform black and coloured schools in matriculation examinations and performance tests at various levels of the school system. Moreover South African educational quality lags far behind even much poorer countries, as has been demonstrated by a number of international tests, including MLA, TIMSS and now SACMEQ II. Educational quality in historically black schools – which constitute 80 per cent of enrolment and are thus central to educational progress – has not improved significantly since political transition.

If it is not possible to indicate that teachers' rights can be positively linked to the quality of education, it seems that the way some of the rights are used or abused amongst others, by unions to usurp management functions may militate against quality. A posting on a blog on the website of the Eastern Cape Education Department states unequivocally that teacher absenteeism leads to lower learner performance.[71] The posting points out that there is national and international statistical evidence that learners whose teachers miss more days of class have lower scores on achievement tests. In other words the learners' learning is disrupted and jeopardised.

Twenty to 24 days are lost per teacher per year in South Africa whereas in the United States teachers are absent from the classroom for an average of 14 days per year.[72] Teachers are entitled to different types of leave and principals must monitor the implementation of legislation in this regard. The most harmful form of teacher absence is of course strikes and protected and unprotected strikes are not uncommon in South Africa.

Furthermore, a media release by the Human Sciences Research Council[73] reports on the gaps between policy and practice regarding the

70 Van der Berg *How effective are poor schools? Poverty and educational outcomes in South Africa* (2006) 3.

71 http://www.ecdoe.gov.za/blog_topic/403/Teacher-absenteeism-leads-to-lower-learner-performance, (accessed 2011-08-15).

72 *Ibid*.

time teachers spend on teaching. An average of 16 hours per week is spent teaching (or 3.2 hours per day) out of an expected range of between 22½ to 27½ hours per week.

In South Africa it would therefore seem that efforts to get teachers (back) in the class for more hours would not be misguided. One way of doing this could be to declare teaching an essential service and to take away or seriously reduce certain labour rights. Having well-qualified and committed and skilful educators in class is a *sine qua non* for the provision of quality education. It seems that teachers' rights to initial teacher education and continuous professional development should be preeminent and could take precedence over rights regarding for example professional control, organisational rights and the right to co-control education (have a voice in education).

Neither South Africa nor Germany provides conclusive evidence of a causal link between teachers' rights and the quality of education. The evidence does invite reflection on assumptions underlying the awarding of labour and other rights to teachers and on views of the nature of teaching. A full prohibition of fundamental rights seems as unacceptable as the full enforcement of fundamental rights without regard to any other considerations.

73 http://www.hsrc.ac.za/Media_Release-254.phtml#mediaRelease (accessed 2011-08-17).

Aantekeninge/Notes

Incoterms® variants: greater precision or more uncertainty?

1 Introduction

The International Chamber of Commerce introduced Incoterms® to achieve some form of international standardisation pertaining to the delivery of goods, passage of risk, allocation of costs and customs formalities under an international contract of sale. These rules "explain a set of three-letter trade terms reflecting business-to-business practice in contracts for the sale of goods" (ICC *Incoterms® 2010* (2010) 5). They are formulated with reference to the most consistent business practices and customs at a given point in time and are regularly updated to keep up with developments in international commercial practice (Foreword to *Incoterms® 2010*). By standardising trade term content Incoterms® achieve legal certainty, which means a reduction in disputes and, hence, also in transaction costs.

The latest revision, Incoterms® 2010, came into operation on 1 January 2011. It introduced a new classification of trade term rules by making a distinction between rules appropriate for all modes of transportation (the EXW, FCA, CPT, CIP, DAT, DAP and DDP rules), and rules aimed exclusively at maritime and waterway transport (FAS, FOB, CFR and CIF). The distinction facilitates the general use of the rules. In addition, the Guidance Note preceding each rule contains advice on the appropriateness of such rule in a particular trade context.

The standardised definitions may, however, be overridden by customs applicable at the place or port where the rule is used (*Incoterms® 2010* 5 par 2). Furthermore, they can be adapted by using Incoterms® variants. Additions to or variations of the basic term can either add to or derogate from the parties' obligations (Raty "Variants on Incoterms (Part 2)" in *Incoterms in Practice* (ed Debattista) (1995) 152-153). For example, an added obligation for the seller to load the goods on the buyer's collecting vehicle in the case of the EXW term, or to pay for costs of discharge in the case of the CIF term. Variants such as "free on board stowed" (FOBS) and "free on board stowed and trimmed" (FOBST) are also regularly used in international trade to extend the seller's delivery obligations.

Whilst Incoterms® merely reflect the commercial practice most commonly used, trade term variants are primarily aimed at obtaining greater precision (ICC *Incoterms 2000* (1999) Introduction par 11; Ramberg *Guide to Incoterms® 2010* (2010) 41; Raty 152-153). Incoterms® recognise the principle of party autonomy and therefore do not prohibit

variations of the standard definitions (*Incoterms® 2010* Introduction 10). However, since there are no consistent practices in regard to the obligations contained under these variants, their content is not standardised and, hence, they create uncertainties. A lack of clarity on the extent of the parties' respective obligations and how far they deviate from the standard rule can create disputes and increase transaction costs. In the case of "EXW loaded" or "FOB stowed", for example, there is no general world-wide consensus that the addition extend the seller's obligations to include both the cost of actually loading the goods and the risk of fortuitous loss of or damage to the goods during the process of loading or stowage. Additions to the C-terms also present many problems. These terms constitute shipment contracts where the seller fulfils his delivery obligations on shipment of the goods. The addition of obligations referring to the destination could suggest that an arrival or destination term was intended. (Ramberg 42-43). In the result, the seller could be at risk until the goods actually arrive at the destination. Divergent opinions are held on whether Incoterms® variants postpone the passage of risk until the additional obligations are performed.

The official rules provide no guidance on how to deal with these variants, except to warn users against the dangers in contracting on this basis (*Incoterms® 2010* Introduction 10). The question that this note addresses is whether Incoterms® variants succeed in providing greater clarity as regards the obligations of the seller and buyer, or whether they actually contribute to more uncertainty. For purposes of this analysis the discussion will be restricted to FOB and CIF variants.

2 Analysis

2 1 FOB Variants

The rationale for the FOB variants has its roots firmly in tradition. International mercantile custom determined that the seller's obligation to deliver the goods on board the vessel nominated by the buyer is fulfilled the moment that the goods cross the ship's rail. Although the ship's rail as a dividing point for risks and costs seems to be a fairly simple solution, it was never appropriate in practice. It is difficult to separate the loading costs for work performed before the point of passing the ship's rail from that performed thereafter, especially if the whole loading operation is conducted by the same company. Moreover, it is also impracticable to divide the functions and risks between the parties while the goods are swinging across the ship's rail. As Devlin J (as he then was) remarked in *Pyrene v Scindia Navigation* (1954 2 QB 402 419):

> Only the most enthusiastic lawyer could watch with satisfaction the spectacle of liabilities shifting uneasily as the cargo sways at the end of a derrick across a notional perpendicular projecting from the ship's rail.

The question of the passing of risk can become especially difficult in the context of accidents during loading operations, for example if the ropes break when the cargo is lifted from the shore onto the ship. One suggestion has been that where the goods are damaged during the

loading process, the risk will be on the seller if they fall on the wharf or in the water, whereas it will be on the buyer if they fall on deck (Valioti *Passing of Risk in international sale contracts: A comparative examination of the rules on risk under the United Nations Convention on Contracts for the International Sale of Goods (Vienna 1980) and Incoterms 2000* (LLM thesis 2003 University of Kent) text accompanying nn 207-212, 266-268). However, because it is purely fortuitous on which side of the rail the cargo drops, this suggestion does not produce a legally satisfying result. In principle, such an accident affects both the contract of sale and the contract of maritime carriage by sea. In the contract of carriage by sea, the loading operation is considered as an indivisible whole and the carrier's liability for negligence extends to all stages of that operation, irrespective of whether they occurred before of after crossing of the ship's rail (Devlin J in *Pyrene v Scindia Navigation supra* 419). However, when it comes to the contract of sale, the situation has never been clear, especially if the parties failed to stipulate this point in their contract.

Generally, there are two views. One view suggests that under an FOB contract the goods are at the buyer's risk when they pass the ship's rail so that it is irrelevant whether they reach the ship safely on completion of the loading operation. The other view suggests that the seller has fulfilled his obligations under an FOB contract only if the goods are deposited safely on board the vessel and the loading operation is completed (Murray et al *Schmitthoff's Export Trade: The Law and Practice of International Trade* (2007) par 2-013).

In practice, merchants developed trade usage variants to deal with the uncertainties that come with the ship's rail criterion in regard to the division of risks and costs. The seller's obligation to place the goods on board may be extended by a phrase added to the FOB term, for example "FOB stowed" (FOBS), "FOB trimmed" (FOBT) or "FOB stowed and trimmed" (FOBST). These variants of the FOB term are aimed at safety, stress and stability of the goods. "Stowing" means ensuring that the cargo is positioned on board the vessel in such a manner as to be safe during the proposed transit, for example to position the cargo so that it is distanced from parts of the ship that generate heat and to ensure the stability or balance of the ship. "Trimming" involves the levelling of the cargo during or shortly after loading so that it is evenly distributed in each hold and throughout the ship as a whole. This process applies to dry bulk cargoes to ensure the stability and structural strength of the vessel. In the case of some cargoes it also ensures that the holds are more efficiently filled or in others, such as in the case of coal, it can reduce the spontaneous heating of the cargo. (Reynolds "Stowing, trimming and their effects on delivery, risk and property in sales 'fobs', 'fobt' and 'fobst'" 1994 (1) *Lloyd's Maritime and Commercial LQ* 119 119-120; Klotz & Barrett *International Sales Agreements: An Annotated Drafting and Negotiating Guide* (1998) 71-72 nn 22, 23). In practice, these duties are performed by stevedores and the costs are included in the freight that is to be paid by the buyer. However, expenses for stowing and trimming can be shifted between buyer and seller. FOBS, FOBT or FOBST indicate

that such costs are shifted to the seller (Reynolds 1994 *Lloyd's Maritime and Commercial LQ* 121; Griffin *Day & Griffin The Law of International Trade* (2003) 58).

Apart from allocating costs, the question is whether these variations have any effect on what constitutes delivery under the contract of sale or on the moment that risk passes from the seller to the buyer. There is no clarity as to whether a seller will only effect delivery once loading, stowing and/or trimming, as the case may be, have been completed, and whether the passage of risk is postponed until these additional obligations are performed.

One opinion is that a resort to trade term variants does not change the point of delivery or the point where risk transfers to the buyer. It only serves to specify the costs for securing and trimming the goods at the port of loading (Raty 155, 161). In the case of an FOB sale where the buyer contracts for carriage, or the seller contracts on behalf of the buyer, the seller is unable to control stowage and trimming as there is no contract between him and those responsible for these obligations. One of the principles on which trade terms are based is that the obligations are kept together, namely that whoever has the custody of the cargo also bears the risk. Deviations of this basic rule distort the obligations and should only take place for special reasons. Moreover, policy considerations of international trade also dictate that risk should follow control (Mikkola "Variants on Incoterms (Part 1)" in *Incoterms in Practice* (ed Debattista) (1995) 144 147-148). Once the goods have crossed the ship's rail and are placed on board the vessel, control is relinquished to the buyer, who will then be protected by marine insurance or the contract of carriage. On strength of these considerations, the stowing and trimming obligations should merely entail an additional financial obligation on the part of the seller and not influence the point of delivery or the passing of risk.

The other view is that the seller's delivery obligation is not met until the stowing and trimming obligations have been completed. According to this view, the law should recognise that parties who insert such clauses into their sale contracts intend to provide for both a physical and a financial obligation, and that their intentions would be frustrated if the term was merely regarded as allocating a financial obligation. Delivery and the passing of risk should therefore be delayed until the stowing and trimming have been completed (Reynolds 1994 *Lloyd's Maritime and Commercial LQ* 121-123, 124-127). (For similar, but qualified, opinions see Klotz & Barrett 72; Treitel "Overseas Sales" in *Benjamin's Sale of Goods* (ed Guest) (2006) par 20-090.) Support for this view is found in an American case, *Minex Shipping v International Trading Company of Virginia and SS Eirini* (1969 303 F Supp 205). In this case, the sales contract provided that bags of cement would be shipped "FOB stowed Polish port". The cement became contaminated during the voyage as a result of soy beans falling on the cement bags. The buyer maintained that the contract term required the seller to sweep, clean and dry the holds of

the vessel. However, the court held that the "stowed" term did not impose these duties upon the seller. The seller was merely obligated to place the cement in the holds "in an orderly, compact manner" and "in such a manner as to protect the goods from friction, bruising, or damage from leakage". The court held that once the cargo was properly stowed, the risk of damage passed to the buyer.

But what is the position after the latest revision of Incoterms®? In pursuance of modern commercial practice, the ICC's standardised FOB rule now omits all references to the ship's rail and states that delivery takes place when the goods are placed "on board the vessel nominated by the buyer at the named port of shipment or by procuring the goods so delivered" (FOB *Incoterms® 2010* article A4; *Incoterms® 2010* Introduction 7). Ambiguities and arbitrary results that have accompanied the notion of the ship's rail are now removed. Because Incoterms® link the passing of risk to the point of delivery, this means that the goods are delivered and the risk passes when they are placed on board the vessel nominated by the buyer (FOB *Incoterms® 2010* article A5). But when are the goods considered to be delivered on board the vessel? The ICC advises that in the absence of port customs and practices between the parties, the default position would be when the goods are "first at rest on deck" in an undamaged condition (see answer to Question 17 "Incoterms Rules Q&A September 2011: ICC general guidance on selected questions on the Incoterms® 2010 rules" available online at http://www.iccwbo.org/Products-and-Services/Trade-facilitation/Incoterms-2010/Q-A-September-2011 (accessed 2012-08-30)). Would that then mean that under FOBS and FOBST delivery takes place and risk passes only when the goods are safely stowed and trimmed? The ICC seems to support this view. In reply to a question on when risk transfers under "stowed and secured/trimmed" variants, the ICC states that the costs for the buyer would "most likely be understood to begin only when the goods were safely stowed/secured/trimmed as set out in the contract and passage of risk would likewise be delayed" (see answer to Question 18 "Incoterms Rules Q&A September 2011 ICC" supra). Port customs should, however, also be taken into consideration when interpreting these variants.

However, this is not necessarily the final word on this issue. The ICC points out that their views "are intended as general interpretive guidance only, and not as authoritative opinion" (Introduction to "Incoterms Rules Q&A September 2011" supra). Furthermore, they choose to state their answer to this particular question in a cautious and general manner. The phrase "most likely to be understood", which introduces the reply, leaves room for other interpretations. This means that there is still no certainty on the legal position concerning delivery and risk under FOB variants. For these reasons, it is important that parties clarify whether the seller is to undertake the responsibilities for loading and stowage operations at his cost or whether this also entails an assumption of risk until the loading and stowage obligations are completed, for example, by adding a phrase such as "FOB stowed, costs and risks in connection with loading on the seller". (See also the second part of the reply to Question 18

"Incoterms Rules Q&A September 2011" supra where the ICC strongly encourages parties to clarify their intentions.)

2 2 CIF Variants

Because the CIF rule may not be suitable for all situations, merchants have sought to tailor the standard-form CIF term to fit specific commercial conditions. For example, if the goods are sold "CIF landed", the unloading costs including lighterage and wharfage charges are borne by the seller or are included into the sea freight which he has to pay. Another example is "CIF undischarged" or "CIF free out" where the intention is that the seller's obligations are limited to those that are to be effected inside the ship's hold in the port of discharge. Costs for unloading should be borne by the buyer (Raty 156; Ramberg 43).

Deviations from the standard-form CIF contract are common where oil is transported by sea. A common modification is the "out-turn" or "landed weight" clause which relieves the buyer from having to pay on the basis of the quantities shown on the bill of lading as is normally the case under CIF. These practices evolved because of the potential for losses during the transportation of oil. A distinction must be made between transportation loss and marine loss. Transportation loss refers to loss in the volume of oil during transit due to evaporation, sludge, accumulation, spillage or measurement error, whilst marine loss covers loss caused by fortuitous events such as vessel destruction, bad weather or war (Lightburn & Nienaber "Out-turn clauses in cif contracts in the oil trade" 1987 *Lloyd's Maritime and Commercial LQ* 177 178).

There is agreement that unavoidable transportation loss is shifted to the seller. However, when it comes to other types of risks, there are conflicting views on their apportionment. According to one view, other types of risk pass on shipment as is the case under the normal CIF rule. The out-turn clause is therefore only a price adjustment mechanism with regard to unavoidable transit losses, leaving the risk of marine loss on the buyer. A more far-reaching approach is that an out-turn clause moves the time at which the risk of loss, whether marine or transportation loss, transfers to the buyer from the point of shipment to the port of destination. This view eliminates the need to distinguish between marine loss and transportation loss (Lightburn & Nienaber 178-179).

Scholars such as Lightburn and Nienaber (179) prefer to deal with the issue as under any normal CIF contract. It is their argument that by choosing CIF as the applicable contract basis, a range of obligations is made part of the contract without the use of express words to that effect. No compelling reasons exist for justifying an exception to the normal CIF rules in the case of an out-turn clause. Traditionally out-turn clauses are a short-hand way of dealing with unavoidable transportation loss as opposed to marine loss. A party seeking to overcome the accepted meaning of an out-turn clause in a particular case should therefore carry the burden of showing that the parties intended to disregard the legal

consequences that would normally flow from its use. If parties wish to deviate from the standard meaning, they should provide for that through express contractual terms (Lightburn & Nienaber 179).

Before the 2004 revision of article 2 of the American Uniform Commercial Code (UCC), section 2-321 UCC reflected a policy decision that the out-turn variant should be construed so as to involve the least possible deviation from the CIF basis, while at the same time taking into account the inevitability of transportation losses. This section dealt with CIF or C&F "Net Landed Weights", "Payment on Arrival", and "Warranty of Condition on Arrival" terms. The Official Comment to the section stated that they are:

> [v]ariations of the C.I.F. contract which have evolved in mercantile practice but are entirely consistent with the basic C.I.F. pattern. Subsections (1) and (2), which provide for a shift to the seller of the risk of quality and weight deterioration during shipment, are designed to conform the law to the best mercantile practice and usage without changing the legal consequences of the C.I.F. or C.&.F. term as to the passing of marine risks to the buyer at the point of shipment.

The UCC drew a clear distinction between "CIF out-turn" contracts and "CIF no arrival-no sale" contracts. Section 2-234 UCC left the risk of loss in "no arrival-no sale" contracts explicitly on the seller. In the case of "CIF out-turn" contracts, section 2-321 split the risk, by placing the risk of necessary loss ("the risk of quality and weight deterioration during shipment") on the seller, and the risk of extraordinary loss ("marine risks") on the buyer. The case law shows that this position correlates with mercantile expectations (Lightburn & Nienaber 180-181). It is therefore likely that the legal position will remain the same even after the provisions have been repealed from the official Code.

There is no statutory provision in the English Sale of Goods Act on CIF out-turn clauses. Although CIF variants may create the impression that the contract becomes a destination or arrival contract, the case law indicates that that they do not affect the character of the contract as a true CIF contract (Murray *et al* par 2-037). In a number of cases where the amount payable was made dependable on the quantity of goods which actually arrive, the courts have held that clauses which shift the risk of loss from the buyer to the seller do not necessarily change the essence of a CIF contract and that it merely entails that the seller should allow a price adjustment after the goods have landed (*Denbigh Cowan & Co v Atcherley* 1921 90 LJKB 836; *The Gabbiano* 1940 P 166 174; *Oleificio Zucchi SpA v Northern Sale Ltd* 1965 2 Lloyd's Rep 496 518; *Oricon Waren & Handels GmbH v Intergraan NV* 1967 2 Lloyd's Rep 83 94).

In *Soon Hua Heng Co Ltd v Glencore* (1996 1 Lloyd's Rep 398), the court held that out-turn or similar clauses are normally intended to relate only to the determination of the price. If the goods are lost and the buyer has already paid an estimated price on tender of the documents, he is not entitled to an adjustment, unless he can prove that the shipped goods

were less in quantity or quality than he has paid for. The out-turn clause should therefore only apply to cases where the weight difference arises from ordinary circumstances. Such a clause does not entail that the whole of the risk is to remain on the seller until actual delivery, since such an interpretation would mean that the contract would no longer be a CIF contract (Treitel par 19-006).

These opinions should be supported inasmuch as they underline the need to preserve the basic character of the CIF term. Variations of the CIF term should not affect the character of CIF as a shipping term where delivery and the passing of risk take place on shipment. An additional obligation does not necessarily, nor automatically, change the risk distribution under Incoterms® (Lightburn & Nienaber 184-185). Risks do not follow from functions and costs. That is evidenced by the C-terms where the seller has to pay for the freight to the indicated destination but does not have to assume the risk of loss of or damage to the goods after dispatch from the country of export (Ramberg 41). CIF variations should therefore merely affect the financial obligations of the parties. If the parties intend to shift the point of delivery or the passing of risk for accidental disasters, they should conclude their contract on the basis of an arrival term.

The out-turn variant, in particular, is intended to cover a range of limited events which would otherwise have fallen under the ambit of the risk rule and should not be used to move the point of delivery or the general point where risk transfers from the seller to the buyer. Moreover, in the case of an out-turn clause, the variation is primarily aimed at a specific type of trade, namely the oil trade, where it reflects the commercial practice of dealing with unavoidable transportation losses. These types of risk have been extensively analysed by industry specialists and are clearly recognised in the oil trade (Lightburn & Nienaber 179). If the parties want to broaden the scope of the out-turn clause they should choose another term or they should state their intention clearly.

3 Conclusion

The discussion and analysis of the FOB and CIF variants have shown that, even though they originate from a need to simplify matters, Incoterms® variants create more uncertainty than clarity (Mikkola 145). Because there is no consistent practice as regards trade term variants, the ICC could not yet standardised their meaning, and thus the organisation cautions against their use (*Incoterms® 2010* Introduction 10).

The ICC has endeavoured to formulate guidelines for interpreting the "stowed" and "stowed and trimmed" variants. However, these formulations do not function as an authoritative opinion and are only intended as general interpretive guidelines. Moreover, because they are expressed in uncertain terms, they leave room for differing interpretations. For these reasons, parties should clarify whether the

seller intends to merely undertake the responsibilities for loading and stowage operations at his cost or whether this also entails an assumption of risk until the loading and stowage obligations are completed.

An additional obligation does not necessarily, nor automatically, change the point that risk passes under Incoterms®. The analysis showed that expressions such as "CIF landed" or "CIF out-turn weights" are normally not interpreted as changing the nature of the CIF term. The word "landed" usually refers to the costs of discharge, and "out-turn weights" merely signifies that the buyer should pay according to the weight ascertained after discharge so that condensation of the goods during their transport should, for instance, be taken into account when fixing the price. However, this does not mean that the seller bears the risk of fortuitous loss of or damage to the goods during their carriage.

Because Incoterms® do not standardise commonly used trade term variants, parties who use them contract at their own peril. Incoterms® leave it either to trade usage in a particular trade sector, to port customs or to the parties themselves to clarify the content of such deviations. If parties wish to deviate from the standard meaning of Incoterms® they should avoid misunderstandings by using express contractual terms which explain their intention clearly (*Incoterms® 2010* Introduction 10). If, for example, the parties merely intend to clarify the extent to which the seller should pay for certain costs without moving the point where risk passes, this should be said explicitly by adding the phrase "discharging costs until placing the goods on the quay for seller's account", or "stowed and trimmed but at the buyer's risk after the goods have been placed on board" to their contract (Ramberg 42-43).

J COETZEE
University of Stellenbosch

Postscript

This note is based on research conducted towards the author's unpublished doctoral dissertation Coetzee *INCOTERMS as a form of standardisation in international sales law: the interplay between mercantile custom and substantive sales law with specific reference to the passing of risk* (LLD dissertation 2010 US).

The winner takes it all! Reflections on the world anti-doping code and the possible criminalisation of doping in sport

"I got caught in Seoul and lost my gold medal. I'm here to try to tell people … it's wrong, to cheat, not to take drugs, they're bad for your health"

(Ben Johnson, 1989 as quoted in Nafziger "International Sports Law" (2004) 147).

1 Introduction

The use of performance enhancing substances in sport has been a phenomenon of time immemorial. The use of performance enhancing substances in order to ensure the best possible results in sports is commonly referred to as doping (see Law "Doping Regulation" in *International Encyclopaedia of Laws* (eds Blanpain & Colucci) (2009) 129-143; Gardiner *et al Sports Law* 3rd ed (2007) 269-291; Gardiner *et al Sports Law* 4th ed (2012) 364-392; Anderson *Modern Sports Law – a textbook* (2010) 113-140; Nafziger *supra* 147-164; O'Leary *Drugs and Doping in Sport – Socio-legal perspectives* (2001) 1-57; 125-146; David *A guide to the World Anti-doping Code – fight for the spirit of Sport* (2008) 1-52; Ioannidis "The application of criminal Law on doping infractions and the 'whereabouts information rule – state regulation v self-regulation" (2010) *ISLJ* at www.buckingham.ac.uk/wp-content/uploads/2010/11/ioannidis.pdf (accessed 2012-12-04)).

Doping in sports is currently a highly debated and controversial issue impacting severely on the spirit of sport and the principle of fair play. The question as to how doping in sport can be effectively combatted poses a major challenge to sport. Recently, doping in sport once again received much attention in the highly controversial Lance Armstrong debacle. Armstrong was banned for life by the United States Anti-doping Agency (USADA) and stripped of his seven Tour de France titles. USADA's comprehensive report pertaining to Armstrong revealed that Armstrong was the centre of a sophisticated doping programme. The report disclosed that Armstrong used, *inter alia*, testosterone and blood transfusions and in addition expected his teammates to likewise take part in using prohibited substances (see "USADA: Armstrong 'cheated' way to top" at http://uk.eurosport.yahoo.com/news/usada-report-proves-armstrong-used-drugs-15393 (accessed 2012-12-04)). USADA's report further provided that Armstrong and his teammates used a range of performance enhancing substances ranging from erythropoietin (EPO), blood transfusions, testosterone, corticosteroids, human growth hormone and masking agents (USADA *supra*). Armstrong subsequently admitted to the use of performance enhancing substances which included, *inter alia*, blood transfusions (See for example the article by

Okuro "Lance Armstrong's doping admission: Cheap confession in a reality of TV culture" *Washington Post* 2013-01-28).

Armstrong's case is yet one more example of doping in sport by a highly acclaimed sport star. A question which inevitably arises upon reading the plethora of newspaper coverage pertaining to Armstrong is whether the sanctioning of doping in sport is by any means effective. Upon closer scrutiny of doping in sport, it becomes apparent that doping encompasses a much wider playing field than merely the athlete doping him or herself. USADA's report on Armstrong confirmed that the doping regime in Armstrong's case comprised of a team of individuals all participating in the doping process.

The World Anti-Doping Agency (WADA) was established on 10 November 1999 with the aims of promoting, coordinating and monitoring the combat against doping in sport (David 1). WADA later sought to develop a standardised approach in respect of the detection and punishment of doping and introduced the World Anti-Doping Code (Code) which was unanimously adopted in March 2003 (David 2-3). The code has been adopted by signatories around the world. The striking reality is, however, that despite WADA's efforts to combat doping and despite the Code, the practice of doping in sport is not decreasing. On the contrary, it seems to be escalating.

The latter inadvertently leads to the question as to whether current anti-doping mechanisms are effective in combatting the phenomenon. It is clear that a more robust approach to doping is essential in eliminating this phenomenon. The question which inevitably falls to be assessed is whether doping in sport should not be criminalised. Ioannidis (1-2) eloquently espouse the following opinion:

> It is submitted that the invocation of such powerful machinery, such as the criminal law, needs to be made with reference to sufficient reasons that can justify its application in the area of doping in sport.

In this contribution I will assess the practice of doping in sport against the backdrop of the instruments currently operative in support of combatting the practice of doping. I will, in addition, address the question as to whether doping should be criminalised.

2 Doping Defined

From the outset it is to be noted that there is no general legal definition of doping. Doping is defined in terms of article 1 of the Code as the occurrence of at least one of the eight anti-doping violations as provided for in article 2 of the Code (Anderson 122). The eight anti-doping violations can be summarised as follows (David 100-131):

(i) Presence of a prohibited substance or its metabolites or markers in an athlete's bodily specimen (article 2.1);

(ii) Use or attempted use of a prohibited substance or a prohibited method (article 2.2);

(iii) Refusing, or failing without compelling justification, to submit to sample collection after notification as authorised in applicable anti-doping rules or otherwise evading sample collection (article 2.3);

(iv) Violation of applicable requirements regarding athlete availability for out of competition testing including failure to provide required whereabouts information and missed tests which are declared based on reasonable rules (article 2.4);

(v) Tampering or attempting to tamper with any part of doping control (article 2.5);

(vi) Possession of prohibited substances and methods (article 2.6);

(vii) Trafficking in any prohibited substance or prohibited method (article 2.7);

(viii) Administration or attempted administration of a prohibited substance or prohibited method to any athlete, or assisting, encouraging, aiding, abetting, covering up or any other type of complicity involving an anti-doping rule violation or any attempted violation (article 2.8).

An in depth analysis of each of the abovementioned violations falls beyond the scope of this contribution. It is important to note that articles 1 and 2 of the Code should be read in conjunction with the preamble to the Code which provides that anti-doping rules equate, in a quasi-contractual fashion to sports competition rules regulating the conditions under which the particular sport is played and athletes are deemed to accept and be bound to the rules as a condition of participation (Anderson 122-123). Articles 1 and 2 should, in addition be read within the context of article 21.1.1 of the Code which provides that athletes are deemed to have constructive notice of their specific roles and responsibilities and "to be knowledgeable of and comply with all applicable anti-doping policies and rules adopted pursuant to the Code" (Anderson 123).

3 Code – Fighting for the Spirit of Fair Play in Sport?

For purposes of the current contribution, emphasis will be placed on the Code as the most prominent international code dealing with the prohibition on doping.

As stated above, the Code was the initiative of the WADA and was the product of a long drafting process. The Code was first adopted in 2003 and became effective in 2004 (Gardiner *et al* 371). The current revised Code came into operation on 1 January 2009 (Gardiner *et al* 371). The main objective of the code was to produce international consensus with the organisations which are signatories to it (David 2). The code contains "core" articles dealing mainly with doping violations, the proof of such violations and sanctions (David 2). The practical implication of the acceptance of the code is that both national and international level athletes who are bound by the code can be subjected to both testing for the presence of prohibited substances in their bodily samples in and out

of competition and to investigation pertaining to various violations which do not necessarily require adverse analytical outcomes or the analysis of bodily samples (David 4). In addition, other individuals who are bound by the code such as athlete support personnel, who may commit anti-doping violations such as trafficking or administering prohibited substances, can also be assessed by anti-doping organisations at both national and international level (David 4). The organisations which can accept the code as signatories save for WADA itself, include the International Olympic Committee, national Olympic committees, the International Paralympic Committee, national Paralympic committees, international federations, national anti-doping organisations and major event organisations (David 3; Gardiner *et al* 371). Neither governments nor national sporting organisations may be signatories to the code (David 3). National sporting organisations are brought within the ambit of the code by means of agreements made by signatories which adopt the code such as international federations and national anti-doping organisations (David 3).

The Code contains a set of regulations that strive to promote consistency in the application of anti-doping regulation (Gardiner *et al* 372). Gardiner *et al* (372) elucidates the problematic aspect of implementing the code as follows:

> One of the major obstacles for any body intent on implementing a rule for the whole world is that it must encompass a diverse range of religious, legal and social perspectives. The imposition at a global level of blood testing, for example would meet with resistance on the grounds of religion and might prove counter-productive for where resistance to the global regime has arisen at a local level, whether as a result of cultural, social or legal disparities, this has had the effect of creating a situation of imbalance, where athletes are either subject to much stricter controls or effectively allowed to act irrespective of the rules 'imposed by the global regulator' (see also Boyes "The International Olympic Committee, transnational doping policy and globalisation" in O'Leary (ed) "Drugs and Doping in Sport" (2000) 178).

The preamble to the Code strives to preserve the essence and intrinsic value of the "spirit of sport" (Gardiner *et al* 372). In terms of the "spirit of sport" various values come into play such as ethics, fair play, honesty, health, excellence in performance, fun and joy, teamwork, dedication and commitment, respect for rules and laws, respect for self and other participants and courage (Gardiner *et al* 372).

Articles 1 and 2 of the Code contain the anti-doping rule violations under the code. The various violations have already been noted under paragraph 2 above. Article 3.1 of the Code provides that the burden of establishing that an anti-doping violation was committed falls on the anti-doping organisation alleging the violation (Gardiner *et al* 377; David 133). The standard of proof will entail whether the particular anti-doping organisation sufficiently established an anti-doping violation to the satisfaction of the hearing panel. The standard of proof is more than a mere balance of probability but less than proof beyond reasonable doubt

(Gardiner *et al* 377). Where the code specifically places the burden of proof upon an athlete or other person alleged to have committed an anti-doping rule violation, the burden of proof is a balance of probabilities except in terms of articles 10.4 and 10.6 where the athlete is required to meet a higher burden of proof (Gardiner *et al* 377; Anderson 128-129; David 133-134). Article 3.2.1 of the Code, in addition, contains a rebuttable presumption that the accredited laboratory conducted the particular assessment correctly (Gardiner *et al* 378).

Article 8 to the Code specifically pertains to the right to a fair hearing. The hearing process should address whether an anti-doping rule violation was indeed committed and, if so, the consequences thereof (Gardiner *et al* 378; David 189).

The hearing process should, *inter alia*, respect the values of a timely hearing; fair and impartial panel; the right to legal representation; the right to be timely informed of the alleged violation and the right to a written and reasoned decision with specific reference to an explanation of the reasons for any period of ineligibility (Gardiner *et al* 378). WADA provides for its own list of banned substances providing not only for a list of substance outlawed, but also their metabolites and 'related' substances (Gardiner *et al* 381). Article 4.3 provides for the following:

> 4.3.1 A substance or method shall be considered for inclusion on the Prohibited List if WADA determines that the substance or method meets any two of the following criteria:
>
> 4.3.1.1 Medical or other scientific evidence, pharmacological effect or experience that the substance or method, alone or in combination with other substances or methods has the potential to enhance or enhances sport performance;
>
> 4.3.1.2 Medical or other scientific evidence, pharmacological effect or experience that the use of the substance or method represents an actual or potential health risk to the Athlete;
>
> 4.3.1.3 WADA's determination that the use of the substance or method violates the spirit of sport described in the Introduction to the Code.
>
> 4.3.2 A substance or method shall also be included on the Prohibited List if WADA determines there is medical or other scientific evidence, pharmacological effect or experience that the substance or method has the potential to mask the Use of other Prohibited Substances or Prohibited Methods.

The sanctioning of doping violations is provided for in articles 9 and 10 of the Code. One of the fundamental cornerstones of the Code is premised on the principle of strict liability. The latter doctrine entails that an athlete is strictly liable for the prohibited substance detected in, and revealed by, the testing of their bodily specimen and that an anti-doping violation is committed regardless of whether the athlete intentionally or unintentionally used a prohibited substance or was negligent in respect thereof (Anderson 123; Gardiner *et al* 387; David 169). Article 9 of the Code reaffirms the position of strict liability in respect of anti-doping

violation and provides that an anti-doping transgression in respect of an in-competition test automatically leads to the disqualification of the individual result in the competition in conjunction with all the resultant consequences such as the forfeiture of any medals or prizes won (David 169). In terms of Article 10, fixed periods of ineligibility are provided for each anti-doping rule violation as well as subsequent violations (David 170). The most important sanctioning provisions can be summarised as follows (Gardiner *et al* 388-389; David 172):

(i) Article 10.2 – the periods of ineligibility for a contravention of article 2.1 (as discussed above in paragraph 2), article 2.2 or 2.6 are for a first violation two years and second violation, life.

(ii) Article 10.3.1 – for violations of article 2.3 or article 2.5 the period shall be two years for a first violation.

(iii) Article 10.3.2 – for violations of articles 2.7 or 2.8, the period of ineligibility shall be four years for a first violation up to a maximum of life.

(iv) Article 10.3.3 – for violations of article 2.4 the period of ineligibility shall be a minimum of one year and a maximum of two years.

(v) Article 10.5.3 – provides that an anti-doping organisation may, prior to a final appellate decision under article 13, suspend a part of a period of ineligibility imposed in an individual case where the athlete or other person provided substantial assistance to an anti-doping organisation or professional disciplinary body which results in the discovery or establishing of an anti-doping rule violation by another person. The extent to which the period of ineligibility may be suspended will depend on the seriousness of the anti-doping rule violation committed as well as the significance of the substantial assistance provided by the athlete or person concerned.

(vi) Article 10.5.4 – provides that a period of ineligibility may be reduced where an athlete or other person voluntarily admits to the commission of an anti-doping rule violation.

(vii) Article 10.10.1 – provides that no athlete or other person who has been declared ineligible may, during the period ineligibility, participate in any capacity in a competition or activity initiated or organized by any signatory, signatory's member organization or a club or other member organization.

In addition article 11 provides for sanctions to be imposed on teams who participate in team sports (Gardiner *et al* 390).

It is clear from the abovementioned summary of the most important principles of Code, that the code provides a well-structured and coordinated regulatory framework aimed at combatting doping in sport and to enhance the spirit of sport. The question which inevitably arises is whether the sanctions imposed upon athletes or other persons who committed anti-doping violations are effective in combatting the phenomenon. Should doping in sport not also be criminalised at a national level? Despite movements in some countries to criminalise doping, doping in sport to a large extent is not regarded as a criminal

offence (see for example Halgreen "The Danish Elite Sports Act" 2005 *ISLR* 74-75; Gardiner *et al* 391).

4 Doping Regulation in South Africa – a Synopsis

In South Africa it is notable that the South African Institute for Drug free Sport (SAIDS) was established in terms of the South African Institute for Drug-free Sport Act 14 of 1997 (hereinafter "the Act"). In terms of the Act, the SAIDS "fulfils an independent testing, education and research function relating to drugs and doping in sport" (Louw 129). In 2006 the Act was amended in order to provide for the enactment of matters contained in the Code and established a doping control programme in line with the code (Louw 129). The definition of doping in terms of the Act corresponds with the definition contained in the Code as well as the anti-doping violations and the burden of proof (Louw 133-136). South Africa is also a signatory to the Code. The SAIDS, in addition, provides for its own unique list of prohibited substances (see www.drugfreesport.org.za (accessed 2012-12-04)).

5 A Move Towards the Criminalisation of Doping in Sport

Despite the immense value of the Code and its regulatory framework in respect of doping, the practice of doping in sport is still a reality. The deterrent effect of the Code in respect of doping remains questionable. A related anomaly pertains to the potential criminal liability of not only the athlete him- or herself, but also of the coaches, medical doctors and others involved in the doping process. Principles of criminal liability which come to fore are *inter alia*, the crime of fraud, aiding and abetting as well as the doctrine of participation. Ioannidis correctly opines that establishing a criminal framework for doping in sport achieves those elements which are currently missing from the sporting governing bodies' regulatory disposition which are certainty, consistency and transparency (Ioannidis (2010) *supra* 15). Ioannidis (15) in addition note that the purpose of a criminal framework:

> [i]s not retribution for an injustice, but the protection of athlete's health, as well as the protection of the social and cultural role of sports, the "fair play" principle, the genuineness of the results and the general and specific prevention.

Criminal sanctions should also pertain to those who encourage and assist an athlete with the use of the prohibited substances. Athletes may perhaps not have the necessary medical knowledge to assess the dosage or the most appropriate time for receiving these substances and will inadvertently turn to doctors or coaches for advice which in turn exacerbates the need for criminalising the conduct of such persons assisting the athlete (Ioannidis (2010) *supra* 16). The regulatory framework currently in place will in all probability be effective in the case of individual transgressions due to the principle of strict liability prevailing, but will be less successful in unveiling the organised crime of doping underlying the system of doping in sport. The main advantages

of criminalising doping are certainty, independent and transparent proceedings and consistence (see "Making doping criminal – the Austrian 'sports fraud' provision and general thoughts on criminalizing doping" in *Sport and the Law* (2010) at http://www.law.ed.ac.uk/courses/blogs/sportsandthelaw/blogentry.aspx?blogentryref = 8 (accessed 2012-12-04) hereinafter referred to as "Making doping criminal"). It could be argued that the aspect of prevention within the criminal context renders the criminalisation of doping more effective to the system currently run by the various sporting governing bodies (Making doping criminal 1). Criminalising doping will, in addition, have the added benefit of public prosecution which entails that doping transgressions will be investigated more effectively due to enhanced legal backing and manpower of the prosecuting authority (Making doping criminal 2). Ioannidis states that doping has an inherent element of cheating and correctly asserts that any manifestation of cheating has the potential to undermine and infringe upon sport (24).

It is submitted that doping in sport should be criminalised in terms of legislation specifically providing for the offence of doping within sport and prescribing penalties for the various violations. Such legislation should obviously be drafted with due regard to the Code and as such the legislation could complement the Code.

Ioannidis (24) mentions the following in respect of the enactment of legislation criminalising doping in sport:

> The adoption of such legislation is intended to reflect the important role sport plays in society and in citizen's lives. In this legislation, the parties involved will accept responsibility for safeguarding the public interest in sport, which encompasses education, professionalism and the ideals of fairness, justice and equality. The legislation will also have to take into account the enormous public interest in sport as a means of promoting health and the vital role that sport plays in improving the health of a nation.

It is submitted that doping impacts severely on the whole principle of sport. It remains an undeniable fact that sport plays an integral and vital role in society. Doping destroys the spirit and purport of sport and will inadvertently result in society losing interest in sport due to the fundamental principle of fair play being diminished. Despite the vital role of WADA in the ultimate struggle against doping, doping violations in sport remains an inescapable reality. It is submitted that a more robust approach is needed to penultimately combat this phenomenon and to protect the reputation and image of sport. The latter can be achieved by means of a statutory framework in terms of which athletes or other persons, who have committed anti-doping transgressions will be prosecuted in terms of the criminal law regime and if found guilty, sentenced according to prescribed statutory sentences. Such legislation should in addition, punish doping on all levels as it is clear that doping is more often than not, a multi-layered programme stretching beyond merely the athlete doping him or herself.

6 Conclusion

Fair play in sport means fairness in all respects. Athletes gaining unfair advantages to other athletes due to the use of performance enhancing substances, need to be prosecuted and punished appropriately as such conduct not only destroys the spirit of sport, but could also prove detrimental to the athlete as well. In addition, all other parties involved in the process of doping should be punished for their involvement in the doping process. A possible way forward would be to criminalise doping in terms of a statutory framework punishing doping on all levels. Legislation criminalising doping can be effectively applied in conjunction with the Code. Public awareness of doping should also be increased with specific reference to the serious implications of doping in modern sport. The detrimental side effects of performance enhancing substances should also be constantly emphasised in an ultimate hope of convincing participants not to use these substances thereby protecting the true spirit of sport and fair play.

G P Stevens
University of Pretoria

Regulated flexibility and the Labour Relations Amendment Bill of 2012

1 Introduction

Contrary to statements by Zwelinzima Vavi, general secretary of the Congress of South African Trade Unions (Cosatu), the South African government is not ignoring a decision adopted at the African National Congress' (ANC's) 2007 national congress in Polokwane to ban "labour brokers" (s 198 Labour Relations Act 66 of 1995 (LRA) refers to "labour brokers" as "temporary employment services" (TES); Anon "Labour Brokers" *Leadership* 19 March 2012 available at http://bit.ly/13Wyg1v (accessed 2013-5-25)). This is confirmed by the ANC's 2009 election manifesto which called for laws that would "ensure decent work . . . introduce laws to regulate contract work, subcontracting and out-sourcing, address the problem of labour broking and prohibit certain abusive practices" ("2009 ANC Election Manifesto" available at http://bit.ly/gP5gKl (accessed 2013-5-25); Benjamin "To regulate or to ban? Controversies over temporary employment services in South Africa and Namibia" in *Labour Law Into the Future: Essays in Honour of D'Arcy du Toit* (eds Malherbe & Sloth-Nielsen) (2012) 189 202). A lot of water has flowed into the sea since these political commitments were made, but all

indications are that the amendments are not far from their implementation.

A package of labour law amendment bills has been published during 2012 and the different pieces of legislation will be implemented in a staggered fashion. The Basic Conditions of Employment Amendment Bill [B15-2012] (BCEAB) is currently before the national assembly, the Labour Relations Amendment Bill [B16-2012] (LRAB) is presently being discussed before a parliamentary portfolio committee and the draft Employment Equity Amendment Bill and a draft Employment Services Bill are yet to be tabled in parliament. Indications are that the BCEAB will be signed into law before the end of 2013 and the LRAB will follow shortly thereafter (Botes & Sishi "Proposed amendments to labour legislation: Where are we?" *SA Labour Guide* available at http://bit.ly/11upyrX (accessed 2013-5-24).

Although a number of issues are covered in the amendments, the package is dominated by the suggested regulation of non-standard forms of work, which include TESs, fixed-term contracts and part-time employees. Three aspects are covered in this contribution: Firstly, what is South Africa's overarching labour policy framework and to what extent have these policies been influenced by the International Labour Organisation's (ILO's) "decent work" agenda and the European Union's (EU's) so-called "flexicurity" policies? Secondly, what are the content and salient characteristics of the suggested amendments pertaining to non-standard work? Thirdly, concluding remarks are made with the emphasis on the question whether an appropriate balance has been struck between the protection of workers' rights and the provision of flexibility in the labour market.

2 Decent Work, Flexicurity and Regulated Flexibility

South Africa was one of the founding members of the ILO in 1919. However, after being criticised for its racial policies during the 1950s and 1960s, South Africa resigned from the ILO in 1964 and was only readmitted shortly before the first democratic elections in 1994 (Van Niekerk *et al Law@work* (2012) 19–20). The ILO sets international norms in the labour environment and prevents unfair competition amongst member states. The implementation of protective measures, such as the principle of equal pay for men and women and social security benefits during the time of unemployment, increases labour costs. One of the purposes of the ILO is to prevent a "race to the bottom" whereby member states may contemplate a reduction of social protection with the view of becoming a more attractive investment proposition (Hepple *Global Laws and Global Trade* (2005) 13).

The ILO functioned optimally in the previous century during an era in which there was a joint commitment towards full employment and providing social security to workers (Hepple 33). This was an era during which the standard eight-hours-per-day, five-day-a-week, indefinite

contract of employment formed the basis of employment relationships. The conventions of the ILO were predominantly rights based providing protection to the weaker party in the employment relationship. Economic and business factors were not the main concern in policy formulation, but rather techniques to provide protection to workers. During the 1950s and the 1960s the eminent labour law scholar, Sir Otto Kahn-Freund, described the key function of labour law as being "a countervailing force" to counter the imbalance in the power relationship between employers and employees (Davies & Freedland *Kahn-Freund's labour and the law* (1983) 18; Le Roux "The purpose of labour law: can it turn green?" in *Labour Law Into the Future: Essays in Honour of D'Arcy du Toit* (eds Malherbe & Sloth-Nielsen) (2012) 230 238).

A lot has changed in the past half a century and the ILO was compelled to reconsider its policy directions. During the post-World War II era the ILO was increasingly being challenged by uneven ratification of conventions, problems regarding the ILO's supervisory mechanisms and globalisation, which saw the exponential growth of non-standard forms of work. As mentioned by Hepple (33), the ILO's response was threefold. It adopted the ILO *Declaration of Fundamental Principles and Rights at Work*, revised and integrated its international labour standards (of which 71 of the 185 conventions were up to date and 54 were outdated) and most significantly for purpose of this discussion, adopted its labour strategy by embracing the "decent work" agenda. The objectives of this policy, which was accepted in 1999, are well known. In a four-pronged approach, it seeks to balance the realisation of fundamental rights at work; the promotion of job creation; the promotion of effective social protection for all; and the encouragement of "tripartism" and social dialogue (Rodgers, Lee, Swepston & Van Daele *The ILO and the Quest of Social Justice 1919-2009* (2009) 205-235). In sum, the decent work agenda has shifted the ILO's attention from a rights-based agenda to one which includes policies that could potentially create jobs and reduce poverty. This sensitivity to labour market conditions is confirmed by the statement of the then Director General of the ILO, Juan Somavia, namely, that "the principle rout out of poverty is work, and to this end the economy must generate opportunities" (ILO *Organising Social Justice,* Report by the Director General (2004) 16).

In 2007 the EU followed suit and changed its labour market policies when it adopted the flexicurity approach (European Commission *Towards Common Principles of Flexicurity: More and Better Jobs through Flexibility and Security* (2007) 359). The term is a combination of the words "flexibility" and "security". Following the Treaty of Lisbon in 2003, a task team headed by the former Dutch prime minister, Wim Kok, devised strategies to counter unemployment and to improve the EU's competitiveness (Kok *Jobs, jobs, jobs – creating more employment in Europe* Employment Task Force Report (2003) 29–30). The underpinning of the strategy is to balance the protection of fundamental rights of workers, but at the same time, to establish flexibility in the labour market to enable employers to respond to changing market conditions. Although

not developed as a one-size-fits-all approach, four broad mutually supportive pathways were conceived which could ultimately provide guidance to member states. The strategy seeks to provide for flexible and reliable contractual arrangements to promote the upward transition of non-standard contractual arrangements to a situation of full protection; to promote investments in comprehensive lifelong learning; to enhance active labour market policies to strengthen the transition of workers between jobs; and to modernise social security systems to enhance the mobility of workers in the labour market (European Commission *Flexicurity Pathways: Turning Hurdles into Steppingstones* (European Expert Group on Flexicurity) (2007) 2-3; Bekker & Wilthagen "Flexicurity – a European approach to labour market policy" 2008 *Intereconomics* 68).

The change of policy direction after the adoption by the ILO of its decent work agenda and the implementation of the EU's flexicurity approach are reflected in the adoption of international standards relating to the regulation of triangular agency work relationships. Rather than prohibiting TESs, the ILO adopted the Private Employment Agencies Convention 1997 (No 181) which seeks to balance flexibility and the protection to agency workers. One of the stated purposes of the Convention is "to allow the operation of private employment agencies as well as the protection of [such] workers" (art 2(3) of the Convention).

The EU has since followed the same policy direction and has implemented the Temporary Agency Work Directive 2008/104/EC, which, amongst others, is aimed at recognising "temporary work agencies", providing protection to agency workers and establishing a framework which could contribute to the creation of jobs (art 2 of the Directive). The Agency Directive is the third of a trilogy of directives which seek to balance the social rights of non-standard employees on the one hand, and on the other, to leave room for flexible working arrangements as part of labour market regulatory strategies. (The other two directives are the Framework Agreement on Part-Time Work Council Directive 97/81/EC and the Framework Agreement on Fixed-term Work 1999/70/EC.):

> To what extent has South Africa adopted a policy framework that seeks to balance the protection of workers' rights without stifling economic growth? The Cheadle Task Team was briefed to prepare South Africa's first set of post-constitutional labour legislation and in its ensuing Explanatory Memorandum it mentioned that the draft Bills sought to "avoid the imposition of rigidities in the labour market" as it aimed to "balance the demands of international competitiveness and the protection of fundamental rights of workers" ("Explanatory Memorandum prepared by the Ministerial Task Team" 1995 *ILJ* 278 285–286).

At more or less the same time an ILO *Country Review Report* swayed influential South African labour law scholars to develop the notion of "regulated flexibility" (Standing, Sender & Weeks *Restructuring the Labour Market: The South African Challenge: An ILO Country Review* (1996)

1–10; Cheadle "Regulated flexibility: Revisiting the LRA and the BCEA" 2006 *ILJ* 663 668 mentions that the "concept of regulated flexibility was developed by Paul Benjamin" based on the *ILO Country Review's* conception of flexibility.) Regulated flexibility amongst others represents a policy framework which provides for the selective application of legislative standards, depending on the remuneration earned by workers and the size of employers' undertakings. Two principles underpin the South African brand of balancing flexibility and the protection of workers' fundamental rights. Firstly, it is recognised that lower earning employees are generally in a more precarious position than higher earning employees, who, through education or experience may have earned a level of security in employability. Secondly, smaller undertakings should not be burdened with obligations that could potentially introduce rigidities and costs which would ultimately inhibit job creation.

The structure of the Basic Conditions of Employment Act 75 of 1997 (BCEA) has until now been the most notable example of the implementation of the regulated flexibility policy (Godfrey & Witten "The BCEA: Statutory, administrative and case law developments" 2008 *ILJ* 2406). Only employees earning below the current threshold amount of R183,006 are entitled to protection in terms of Chapter II of the BCEA which covers aspects like maximum hours of work (45 hours in any week; s 9(1)(a) BCEA) and maximum overtime (10 hours in any week; s 10(1)(b) BCEA). In similar vein, only employees earning below the threshold amount can rely on the rebuttable presumption regarding who is deemed to be an employee (s 83A BCEA; s 200A LRA). In respect of the size of undertakings, the Employment Equity Act 55 of 1998 (EEA) provides that only employers who employ fewer than 50 workers fall within the definition of "designated employer" (s 1 EEA).

It is argued that South African policy makers do take account of the fact that it is not the sole purpose of labour law to provide protection to workers. At least some thought goes into the notion that different categories of workers need different levels of protection and that start-up undertakings should not be burdened by regulations to the same extent as larger undertakings. However, as pointed out by Godfrey and Witten (*op cit* 2408-2409), attempts by policy makers to introduce the policy of regulated flexibility have been tempered by the strong position adopted by trade unions in South Africa. The authors point out that the final 2002 amendments resulted in "some balance being restored to the combination of flexibility and 'core' conditions" and turned out to be a "significant victory of labour".

Added to this, it is clear that South Africa has developed its own brand of balancing flexibility and regulation. As part of an integrated labour law strategy the ILO and the EU also emphasises the improvement of social dialogue between social partners, investment in education and training and the modernisation of social security systems. Yes, in South Africa there has been a dramatic improvement of the level of social security

protection that is being extended to the needy, and yes, investments are being made towards skills development, but it seems that there is ample room for improvement regarding the integration and harmonisation of these strategies into a coherent labour policy framework (Benjamin "Labour law beyond employment" in *Reinventing Labour Law: Reflecting on the First 15 Years of the Labour Relations Act and Future Challenges* (eds Le Roux & Rycroft) (2012) 21 35 confirms that expenditure on social grants have increased from approximately R30.1 billion in 2001/2002 to R101.4 billion in 2008/2009)).

3 LR Amendment Bill 2012: Protection of Non-standard Workers

3 1 Introduction

Although most of the media attention has focused on those amendments dealing with TESs, the thrust of the amendments is broader in so far as they cover three categories of non-standard employees (see eg Mawson "Labour broking to be regulated, not banned" (9-7-2012) available at http://bit.ly/171UbaE (accessed 2013525)). As will be discussed below, apart from employees placed by TESs, fixed-term and part-time employees are also provided with improved protection in terms of the suggested amendments. In the EU agency work was valued as potential "steppingstones" for job applicants to enter the workplace before the Temporary Agency Work Directive was adopted in 2008 (Kok *op cit* 29–30 mentions that temporary agency work was impaired by "legal obstacles" in the EU and they suggested that the "removing of obstacles . . . could significantly support job opportunities and job matching"). The discussion below considers the extent to which the suggested amendments seek to establish a balance between protection and regulation in South Africa and the extent to which the policy of regulated flexibility has influenced this round of proposed amendments.

3 2 TES

The current LRA provides that the TES is the employer in instances where (a) a TES procures the services of an employee; (b) the TES remunerates the employee; and (c) the employee renders services to a client (s 198(2) LRA). Furthermore, the LRA confirms that the TES and the client are jointly and severally liable in respect of obligations established by a collective agreement concluded in a bargaining council, binding arbitration awards that regulate conditions of employment and the provisions of the BCEA (s 198(4) LRA.) Despite the fact that the word "temporary" forms part of the term TES, the LRA in its current format is silent regarding the duration of these triangular relationships. The LRA also does not extend joint liability in respect of unfair dismissal and unfair labour practices between the TES and the client. There is also no obligation on TESs to provide agency workers with equal conditions of service (especially equal pay for similar work) compared to workers who are in the employ of clients doing essentially the same work. This has

resulted in a situation in terms of which agency workers are being exploited. In the *Regulatory Impact Assessment Report*, which preceded the set of amendments it is stated that:

> There are also documented cases of large employers employing their entire workforces through TES. Reported case law includes instances of employers 'transferring' their employees to TES, and employees who are unaware that their employer is in fact a TES ... Section 198 can be used by employers to deprive employees of protection against unfair dismissals ... and to apply less favourable terms and conditions of employment
> (Benjamin, Bhorat and Van der Westhuizen "Regulatory impact assessment of selected provisions of the: Labour Relations Amendment Bill, 2010, Employment Equity Amendment Bill, 2010, Employment Service Bill, 2010" available at http://bit.ly/1au4fpO (accessed 2013-5-27) 32).

The new section 198A of the LRAB suggests that there will be improved protection afforded to agency employees who earn below the earnings threshold in terms of the BCEA which currently stands at R183,008 per annum. In other words, the same level of protection (some would say lack of protection) applies to all employees earning above the threshold amount. Three categories of protection will be introduced:

(1) Temporary in nature: The LRAB provides that employees not rendering "temporary services" will be "deemed to be the employee of the client and the client will be deemed to be the employer" (s 198A(3)(b)). Furthermore, "temporary service" is defined to mean work not exceeding a period of six months, work performed as a substitute for an employee of the client who is temporarily absent or work that has been categorised as such by a collective agreement or a sectoral determination (s 198A(1)(a)-(c)).

(2) Termination to avoid consequences: The LRAB proposes that should a TES terminate the assignment of a worker to a client in order to avoid the operation of the section that deems the worker to be an employee of the client, the termination will be deemed to be a "dismissal" in terms of section 186(1) of the LRA.

(3) Equal treatment: The LRAB suggests that unless there is a justifiable reason to do so, an employee deemed to be the employee of a client must on the whole be treated "not less favourably than an employee of the client" doing similar work (s 198A(5)).

When analysing the suggested protective measures the following aspects need to be highlighted. It is predicted that the words "deemed to be the employer" will cause uncertainty. Would this mean that the client does not become the actual employer and that the TES remains a party to the contract of employment after the initial six-month period? The provision only "deems" the client to be the employer, but does not stipulate that the client steps into the shoes of the TES in respect of the contract of employment. This is the statutory position in respect of transfers of businesses as going concerns (s 197 LRA). Even though it is clear that the client will bear the responsibilities of an employer, especially in respect of unfair dismissal and unfair labour practices, the suggested provision does not make it clear that the TES will be absolved of responsibilities in

terms of the initial contract of employment. It is suggested that rather than deeming the employee to be the worker of the client, the section should rather have maintained the position that the TES is the employer with the added protection of rendering the TES and the client jointly and severally liable for all employer-related obligations after the initial six-month period. This, with the equal treatment provision, would prevent the interpretational problems that will in all likelihood result from the suggested amendments.

In respect of the second protective measure, it can be foreseen that TESs will probably rely on the provisions of section 189 of the LRA and subject affected employees to dismissal on grounds of operational requirements to prevent the employee from becoming an employee of the client. If the drafters of the amendments had been serious about blocking such terminations it would have been more appropriate to classify them as "automatic unfair" dismissals, which attract the more stern sanction of a maximum of 24 months' compensation (ss 187(g), 194(3) LRA). This is the way in which dismissals associated with the transfer of businesses as going concerns are currently dealt with (s 187(1)(g)).

Taking a leaf from the EU Temporary Agency Work Directive, it seems like an omission in the amendments in so far as no provision is being made for any guarantees for agency employees regarding the right to be informed of, and the right to apply for, vacant positions at the client. It could also have been useful had the amendments suggested that any agreement which has the effect of preventing agency employees from concluding contracts with a client after an assignment will be declared void (art 6 of the Temporary Agency Work Directive). This principle can be linked to the sentiment that agency work can serve as a potential spring board for job seekers into the labour market.

3 3 Fixed-term Contracts

The current provisions of the LRA provide limited protection to employees engaged in fixed-term contracts. So, for example, there is no protection in respect of the maximum duration of fixed-term contracts, or any obligation to apply the principle of equal pay for equal work performed. The only mentionable protection is in respect of the non-renewal of fixed-term contracts. The LRA presently provides that in instances where an employee "reasonably expected" the renewal, and where the employer does not offer to renew a fixed-term contract (or only offers to renew the contract on less favourable terms), it constitutes dismissal (s 186(1)(b) LRA). From this it follows that once it is determined that it is a dismissal, the employer bears the onus to prove that there was a sound reason for the dismissal and that a fair pre-dismissal procedure had been followed (s 188 LRA). The Labour Appeal Court has interpreted the current section to mean that it does not protect employees who reject reappointment on a fixed-term contract on grounds that they may have expected to be appointed on a permanent basis pursuant to consecutive

fixed-term contracts (*University of Pretoria v Commission for Conciliation, Mediation and Arbitration* 2012 *ILJ* 183 (LAC)). This, the amendments suggest, should be changed. A new section 186(1)(b)(ii) will read that "dismissal" means that:

> (b) an employee engaged in a fixed-term contract of employment reasonably expected the employer -
>
> (ii) to retain the employee on an indefinite contract of employment but otherwise on the same or similar terms as the fixed term contract, but the employer offered to retain the employee on less favourable terms, or did not offer to retain the employee.

This is an improvement on the current situation. In practice, and if based on supporting evidence, it is possible that consecutive contracts can play a role in establishing an expectation in respect of the indefinite appointment of an employee rather than merely being offered another fixed-term contract. However, it can be argued that the wording of the suggested inclusion is vague and could be interpreted to have two meanings. The clause can be reduced for ease of reading to say that "an employee ... reasonably expected the employer ... to retain the employee on an indefinite contract of employment but otherwise on the same or similar terms as the fixed term contract, but ... did not offer to retain the employee". The first interpretation is the one that the drafters are aiming at, namely, to cover a situation where the previously fixed-term employee is not offered an indefinite contract. However, the clause does not say as much. It only says "but ... did not offer to retain the employee". This leaves room for a second interpretation, namely, that the employer did not offer a further fixed-term contract on the same or better terms. This problem could have been solved by adding towards the end the provision "but ... did not offer to retain the employee on an indefinite contract".

Apart from the abovementioned protection offered to fixed-term employees, a new section 198B introduces significant additional protection for employees who earn below the current earnings threshold of R183,008. A "fixed term contract" of employment is defined to mean a contract that terminates on "the occurrence of a specified event", "the completion of a specified task" or "a fixed date other than an employee's normal or agreed retirement age" (s 198B(1)(a)-(c)). Fixed-term employees will be protected in a number of key ways:

> (1) Maximum duration: The LRAB suggests that, unless it can be justified, an employer may only engage an employee on a fixed-term contract (or successive fixed-term contracts) for a maximum duration of six months (s 198B(3)). An extensive list of justifications has been enumerated to which I shall return later. Should an employer without justification engage an employee on one or more fixed-term contracts for a longer duration than the mentioned six-month period, the contract of employment is "deemed to be of an indefinite duration" (s 198B(5)).
>
> (2) Written offer: The amendments propose that an offer to employ an employee on a fixed term, or an offer to extend a fixed term, must be

in writing and it must state the reason justifying the fixed-term appointment (s 198B(6)(a)-(b)).

(3) Equal treatment: The LRAB puts forward that a fixed-term employee employed for longer than six months must be treated "on the whole not less favourably" than indefinitely employed workers performing similar work, unless there is justification for dissimilar treatment (s 198B(8)).

(4) Application for vacancies: An employer must provide fixed-term employees with the same opportunities to apply for vacancies as it applies to employees employed on an indefinite basis (s 198B(9)).

(5) Severance pay: Employees engaged for justifiable reasons for a longer duration than 24 months must be paid one week's remuneration for each completed year in accordance with the provisions of the BCEA (s 198B(10)).

Of significance for purposes of this discussion are the exclusions to the mentioned protection and the justifications. It is clear that the drafters were mindful of the fact that start-up and smaller enterprises may find it difficult to comply with the amendments' additional protection for part-time employees. In what can be classified as one of the most significant expressions of the policy of regulated flexibility, the amendments suggest that the protective measures do not apply to employers employing less than 10 employees, or employers whose business has been in operation for less than two years and that employ less than 50 workers. This exception does not apply in respect of employers with less than 50 workers who conduct more than one business or if the business was formed by the division of a business (s 198B(2)(a)-(b)).

Also of importance is the fact that the reasons which have been included in the amendments, which could justify the appointment of fixed-term employees beyond six months, are ample and cast in wide terms. The reasons include replacing an employee who is temporarily absent; being a recent graduate or student who is being employed for training; is engaged exclusively to work on a "genuine and specified project"; being engaged in seasonal work; employees who have reached the normal or agreed retirement age; and employees who have been engaged for a trial period of less than six months for the purpose of determining whether the employee is suitable for employment (s 198B(4)).

Two aspects need to be highlighted – the first is positive and the second raises a concern. The drafters have adopted a constructive approach by limiting the use of fixed-term contracts to six months in situations where there is no justification for not employing such workers in terms of an indefinite contract. Furthermore, rather than merely providing that it must be justified, specific reasons have been provided which could have the result of limiting the need to litigate in order to establish the boundaries of justifiable reasons. In addition, the list of reasons which could justify longer fixed-term contracts is not closed. The door is still open to "demonstrate any other justifiable reason" of fixing a longer term for a contract of employment (s 198B(3)(b)). An aspect that

is worrying though is that one of the reasons being mooted as being justifiable is to appoint an employee on a fixed-term contract rather than following the route of placing an employee on probation. Although this will provide employers with flexibility in the form of an escape clause in respect of newly appointed employees, it may have the consequence of undermining the established guidelines contained in the Code of Good Practice: Dismissal, which deal with probation and dismissal on grounds of poor work performance (Its 8, 9 Sch 8 LRA).

3 4 Part-time Workers

The LRA presently provides no specific protection to part-time employees. The amendments suggest a definition for both a "part-time" and "comparable full-time" employee. A part-time employee means an employee who is paid by reference "to the time that the employee works and who works less hours than a comparable full-time employee". A "comparable full-time employee" is defined as an employee who is "remunerated wholly or partly by reference to time that the employee works" and who "in terms of custom and practice" of the employer is identifiable as a full-time employee (s 198C(1)(a)-(b)).

The limitations and protections in respect of part-time and fixed-term employees correspond in a number of respects but there are also a number of differences. The protective measures for part-time employees do not protect employees who earn in excess of the threshold amount and employers with less than 10 employees, and start-up employers with less than 50 workers are excluded (s 198C(2)(a)-(b)). Apart from these exclusions, the provisions also do not apply to employees who ordinarily work for an employer less than 24 hours a month and during an employee's first six months of employment (s 198C(2)(c)-(d)).

The main protective measures are that part-time employees who earn below the threshold (a) are entitled to be treated on the whole not less favourably than comparable full-time employees doing similar work unless there is a justifiable reason for different treatment; (b) such workers should be provided access to skills development and training; and (c) they should receive the same access to opportunities to apply for vacancies as full-time employees (s 198C(3)-(4)).

The main concern lies within the definition of part-time and comparable full-time employee. The definition states that a "comparable full-time employee" is one who is "remunerated wholly or partly by reference to time that the employee works" and a "part-time employee" is a person who is paid by reference "to the time" that the employee works. Does this mean that it only applies to persons who are being paid on an hourly basis or does it also refer to persons who are being remunerated on a monthly basis? One could arguably have a half-day and full-day secretary at the same workplace who are being remunerated on a monthly basis and not necessarily for the "time the employee works". Yes, such employees will generally agree to five or six working

days a week, and there will be agreement on whether they either work between 08:00 and 17:00, or between 08:00 and 13:00. However, such employees are paid their agreed upon salary at the end of each month irrespective of the fact that February and March generally do not have the same number of working days. Does this mean that they do not fall under the mentioned definitions as they are not remunerated specifically by reference "to the time" that they work? It is suggested that this problem could have been avoided by omitting the words "time the employee works" from both definitions.

4 Concluding Remarks

At more or less the same time when the flexicurity approach was adopted in the EU, the South African Government briefed Halton Cheadle in 2006 to consider the conceptual underpinnings of the local policy of regulated flexibility as it was applied to the post-constitutional labour laws introduced in the mid-1990s. In a sobering report, Cheadle points out that despite the adoption of the policy of regulated flexibility, the phased and urgent nature of the reforms after the first democratic elections resulted in piecemeal negotiations about specific statutes and it prevented the formation of a logical integrated set of labour laws ("Regulated flexibility" *op cit* 663). This sentiment is echoed by Benjamin, who in 2012 stated that the problems of unemployment and appropriate regulation "would best be dealt with in a broad framework covering the range of labour market regulation. Unless this can be achieved, the terms of the debate will remain too narrow" ("Labour law beyond employment" *op cit* 40).

It is clear that the drafters of the amendments under discussion were mindful of the overarching policy of regulated flexibility and that they did not merely introduce new obligations on employers without taking account of the fact that it may impact negatively on especially smaller employers and lower earning employees to implement additional obligations on employers. Furthermore, it is clear that some flexibility remains in place for employers, in so far as all of the added protective measures only become effective after six months after the employment of non-standard employees.

South Africa can gain direction from both the "decent work" and "flexicurity" policy approaches adopted by the ILO and the EU respectively. In both instances their policies seek to balance the protection of employees' rights with aspects like investment in training and social security measures and the promotion of social dialogue. The purpose of labour law should be extended beyond the mere fortification of employees' rights. This field of study should be alert to aspects that may discourage (and encourage) job creation and which integrates skills development and social security protection into a nuanced and harmonised labour market policy. As pointed out, although the suggested amendments are not perfect in all respects, they were definitely influenced by the principles of regulated flexibility which seek to strike a

balance between providing protection to employees and balancing this with some flexibility for employers. At the very least, it is a positive development that the TES industry was not prohibited and that the imposition of enhanced protection for fixed-term and part-time employees were not extended to small employers and start-up businesses.

B P S VAN ECK
University of Pretoria

The consequences of pleading a non-admission

1 Introduction

The purpose of pleadings is to define the issues in dispute in a civil case, not only for the judge but also for the other party. The opponent must be properly informed of the case he has to meet (*Hillman Brothers Ltd v Kelly and Hingle* 1926 WLD 153). Consequently a party has a duty to allege in his pleadings the material facts upon which he relies (*Minister of Safety and Security v Slabbert* 2010 2 All SA 474 (SCA) 475).

It is unfair to ambush one's opponent at trial by facing him with a case different to the one presented in the pleadings. Rule 22(2) of the Uniform Rules of Court provides:

> The defendant shall in his plea either admit or deny or confess and avoid all the material facts alleged in the combined summons or declaration or state which of the said facts are not admitted and to what extent, and shall clearly and concisely state all material facts upon which he relies.

If a party has no knowledge of assertions made by his opponent he is not in a position to either admit or deny the averments. Hence he may plead that he does not admit certain facts. If a party pleads in this manner he must in terms of rule 22(2) of the Uniform Rules of Court "clearly and concisely state all material facts upon which he relies."

There seems to be a measure of consensus by the courts regarding what constitutes a technically correct pleading when the pleader does not admit an averment. However, there is more controversy concerning what such a pleader may and may not do at the trial. The purpose of this note is twofold:

(i) To ascertain how one must plead in a situation where the pleader does not admit an averment (as opposed to denying an averment) in order for the plea to be technically correct; and

(ii) to determine whether the pleader in these circumstances is entitled to -

(a) cross examine his opponent and his witnesses on the issues that are not admitted; and

(b) call and examine one's own witnesses on the issues that are not admitted.

It is our view that the questions as to technical correctness of the pleading of non admission and the question whether such a pleader may rebut his opponent's averment by calling witnesses, are questions that are inextricably linked. The reason for this is that the manner of pleading must inform the opponent of the case he will have to meet. If the pleader adequately informs his opponent of the case he faces, this will be a technically correct pleading. Furthermore, if the plea is technically correct, the pleader should be entitled to cross examine his opponent and any witnesses his opponent may call. Whether the pleader can call his own witnesses (other than expert witnesses) is questionable.

The conclusions reached in this note are based on the premise that an important purpose of pleadings is to clearly inform one's opponent of the case he has to meet.

2 The Purpose of Pleadings

Pleadings serve the purpose of identifying the issues in a particular matter. The judge refers to pleadings in the course of a trial for the purpose of *inter alia* deciding whether evidence is admissible. Evidence that is extraneous to the issues identified in the pleadings is not relevant and therefore inadmissible. Pleadings:

> [m]ust ensure that both parties know what the points of issue between them are, so that each party knows what case he has to meet. He or she can thus prepare for trial knowing what evidence he or she requires to support his own case and to meet that of his opponent.
> (*Becks Theory and Principles of Pleadings in Civil Actions* (2002) 44).

In *Minister of safety and Security v Slabbert* (2010 2 All SA 474 (SCA) 478 par 11) Mhlantla JA stated:

> The purpose of pleadings is to define the issues for the other party and the court. A party has a duty to allege in the pleadings the material facts upon which it relies. It is impermissible for a plaintiff to plead a particular case and seek to establish a different case at trial. It is equally not permissible for a trial court to have recourse to issues falling outside the pleadings when deciding a case.
> (See *Nieuwoudt v Joubert* 1988 2 All SA 189 (SE) 194 where the court simply stated that the purpose of pleadings is to define the issues and inform one's opponent of what case he has to meet).

In order to determine whether a pleading is technically correct, the question to be posed is whether the pleading in question properly defines the issues so that the other side is in a position to ascertain what evidence it requires to pursue its case at the trial.

3 Technically Correct Pleadings

In order to establish how to plead a technically correct non-admission, it is necessary to understand how and why the Rules of Court were expanded so as to allow a party to plead a non admission. Marais AJ *in Standard Bank Factors Ltd v Furncor Agencies (Pty) Ltd* (1985 (3) SA 410 (C) 417) explains the evolution of the Rules of Court to include the possibility of a non-admission as follows:

> The Rules of Court relating to pleading usually confined the defendant to one of three possible courses. He could admit, or deny, or confess and avoid the plaintiff's allegations. No specific provision was made for a non-admission (as opposed to a denial). Moreover, the Rules then in force provided that 'every allegation of fact not specifically denied in the plea shall be taken to be admitted.' The net result of all this was that a defendant who genuinely had no knowledge as to whether a particular allegation made by plaintiff was correct, was compelled to deny it in order to avoid being taken to have admitted it. It soon came to be recognised that the specific options for which the relevant Rule of Court provided were too limited and did not cater for a case which frequently arises in practice, namely the case where the defendant cannot admit an allegation because he has no way of knowing whether or not it is correct, but on the other hand, does not wish to deny it positively and so create the impression that he intends to contradict it at the trial. A practice developed of allowing the defendant, notwithstanding the silence of the Rules in this regard, to plead non-admissions instead of denials, and of regarding such a plea as sufficient to oust the presumed admission which the Rules provided should be the consequence of a failure to enter a specific denial.

Rule 22(2) of the Uniform Rules of Court now specifically provides for the option of pleading a non-admission. From this it follows that there is a distinct difference between a non-admission and a clear denial. (See *Wilson v South African Railways and Harbours* 1981 3 SA 1016 (C) 1018; *Standard Bank Factors Ltd v Furncor Agencies (Pty) Ltd* 1985 (3) SA 410 (C) 417). This distinction is necessary because of the difference in the practical effect of the respective pleadings. The view taken by Van den Heever in *N Goodwin Design (Pty Ltd v Moscak* (1992 1 SA 154 (C) 163) that the only difference between a non-admission and a clear denial is merely a matter of emphasis, begs the question as to the purpose of the rules providing for both a denial and a non-admission. This is especially so given that the option of pleading a non-admission was specifically and intentionally added to the Rules after the practice of pleading a non-admission had developed in order to cater for situations where the pleader had no knowledge of a particular allegation and therefore was neither in a position to admit or deny the allegation. More fundamentally the view of Van den Heever J ignores the basis for the distinction, namely that the purpose of pleadings is to properly inform one's opponent of the case he has to meet.

The question as to whether a reason for the non-admission must be pleaded arose in various cases (*Wilson v South African Railways and Harbours* 1981 3 SA 1016 (C); *Standard Bank Factors Ltd v Furncor*

Agencies (Pty) Ltd 1985 3 SA 410 (C); *Nqupe v MEC, Department of Health & Welfare, Eastern Cape Province* 2006 JOL 16933 (SE)). In *Wilson* the court held (1018) that rule 22(2) must be read in conformity with existing practice. Therefore the court found that it is not sufficient to plead a non-admission without more. Furthermore, the court held that if it were permissible to plead a non-admission without more, there would be no reason for the rule 22(2) to provide that a pleader must admit, deny or confess and avoid. Furthermore, the court held that a plaintiff is entitled to know what the defendant's defence is and that a bare denial does not properly state what one's defence is. In *Standard Bank Factors Ltd* counsel for the plaintiff argued that since rule 22(2) requires a pleader to "clearly and concisely state all material facts upon which he relies", a pleader may only plead a non-admission on the basis of no knowledge when the facts warrant the lack of knowledge. Counsel for the plaintiff argued that where the defendant's lack of knowledge is self evident it is not necessary for the defendant to explain the reasons for the lack of knowledge. However, counsel reasoned, that in situations where it would appear *prima facie* that the defendant should have knowledge, it is incumbent upon the defendant to divulge the reasons for the lack of knowledge. Marais AJ rejected this argument (416) on the basis that it conflates technical legitimacy and ethical legitimacy. Marais AJ held that ethical legitimacy can rarely be tested at exception stage as evidence at trial stage would be necessary to determine the *bona fides* of a non-admission based on lack of knowledge. Marias AJ therefore concluded that a non-admission which does not explain why the pleader has no knowledge is technically sound even in situations where it would *prima facie* appear that the plaintiff should have such knowledge. Marais JA held that the *Wilson* case is not authority for the proposition that a pleader must explain why he has no knowledge. It merely is authority that the pleader must state that the reason for the non-admission is a lack of knowledge. Marais AJ left the question open as to whether the court in *Wilson* was correct in its findings. The court in *Nqupe*, on the other hand, seems to agree with the arguments put forward by the plaintiff's counsel in *Standard Bank Factors Ltd. The* court held that in circumstances where the defendant should have knowledge of the facts, the pleadings should provide reasons for the lack of knowledge. In this case the defendant (employer) admitted that the plaintiff was employed by it, but pleaded a non admission, based on a lack of knowledge with regard to plaintiff's post and remuneration. The court concluded (16933) that in these circumstances, it would be incumbent on the pleader, not only to state that the basis for the non-admission is a lack of knowledge, but also to explain the reason(s) for the lack of knowledge. A failure to do so, the court held, results in the pleading being technically inadequate since the requirement in rule 22(2) that the non-admission must state clearly and concisely all material facts upon which the pleader relies would not be met. In summary the courts seem to be in agreement regarding the necessity to state the reason for the non admission. This is usually a lack of knowledge. However, there seems to be less consensus with regard to whether the pleader who pleads a non admission is obliged to not only

state the reason therefore (usually lack of knowledge), but to also state or explain the reason for the lack of knowledge. If there is an obligation on the pleader to state the reasons for the lack of knowledge there is no clarity from the case law whether or not this is necessary only in circumstances where it would appear *prima facie* that the pleader should have knowledge.

We submit that all these questions must be answered with reference to the most important purpose of pleadings, namely to inform one's opponent of the case he has to face. So for example, if a lack of explanation or reason for the lack of knowledge would result in one's opponent not knowing the case he faces, then in order for the pleading to be technically correct, it should state the reason for the lack of knowledge.

4 Cross-examination of Opponent's Witnesses and Leading of Evidence in Rebuttal

In *Goodwin Design (Pty) Ltd v Moscak* (1992 1 SA 154 (C) 163) Van den Heever J expressed the view *obiter* that the following statement of Bullen, Leake and Jacobs (*Precedents of Pleadings* (1975) 80) is a correct reflection of the law: "there is no difference in effect between denying and not admitting an allegation. The distinction is simply a matter of emphasis, a denial being more emphatic than a non admission." Consequently, Van den Heever J concluded that a defendant who pleads a non-admission is entitled not only to cross examine plaintiff and his witnesses but also to call his own witnesses and lead rebutting evidence. She added that in the event that the defendant's witnesses are expert witnesses, the plaintiff would be given notice that they would be called in terms of the Court Rules and would therefore not be caught by surprise. Van den Heever J justifies her conclusion by arguing that if the pleader of a non-admission is not permitted to cross-examine the plaintiff or adduce evidence, the plaintiff's version would prevail by the plaintiff simply re-iterating what was stated in the pleading at the trial. This is not entirely correct. This would only be the case in situations where the court could be convinced that the plaintiff's evidence was sufficient to satisfy the plaintiff's onus of proof (Zeffert & Paizes *The South African Law of Evidence* (2009) 129-130). In order to do this the plaintiff must adduce sufficient and credible evidence to establish a *prima facie* case. A failure to do so will result in the court not accepting the plaintiff''s version even in the absence of evidence in rebuttal from the defendant (Zeffert & Paizes 132).

The view of Van den Heever J was not shared by the courts in other cases (*Ntshokomo v Peddie Stores* 1942 EDL 289; *Wilson v South African Railways and Harbours* 1981 3 SA 1016 (C) and *Standard Factors Ltd v Furncor Agencies (Pty) Ltd* 1985 3 SA 410 (C)). The reason for this was that a plaintiff faced with a positive denial can expect the defendant to lead rebutting evidence whereas a plaintiff faced with a non-admission need not, hence the difference between a non-admission and a denial

(*Standard Factors Ltd v Furncor Agencies (Pty) Ltd* 410). The court however held that it may be going too far to contend that the defendant "need not even anticipate a challenge by way of cross-examination". In *Ntshokomo v Peddie Stores* (289), on the other hand the court held that that the pleader of a non-admission is not entitled to lead evidence or to cross-examine the plaintiff on matters that were not admitted.

The question whether a party should be allowed to cross-examine his opponent or his opponent's witness regarding a matter to which he pleaded a non-admission is dependent upon the purpose of cross-examination. Cross-examination has two purposes, "first, to elicit evidence which supports the cross –examiner's case, and second, to cast doubt upon the evidence given for the opposing party." (Zeffert & Paizes 909). It is our submission that even in circumstances where the cross-examination relates to issues where a non-admission on the basis of lack of knowledge was pleaded, the pleader should be allowed to cross examine the opponent's witness. Knowledge of facts is not necessary for cross-examination. If for example, the evidence-in-chief of the witness was flimsy or contradictory, cross-examination may cast further doubt on the credibility of the witness or the evidence without the necessity of calling one's own witness to contradict or cast doubt on the evidence. One should always be given the opportunity to cross examine an opponent or his witnesses, even if one has no knowledge of events. This is to test credibility of the witness and of the version put forward by the witness.

However, if the pleader of a non-admission based on a lack of knowledge were allowed to lead evidence to contradict the plaintiff's version, this would imply that the defendant in fact had knowledge. It is doubted that an order for costs against such a pleader, would be sufficient to remedy the potential injustice suffered by a plaintiff if the defendant were allowed to lead evidence (that is not expert evidence) in rebuttal of the plaintiff's case. This is because the plaintiff now faces a case for which he was not prepared. This runs contrary to the fundamental purpose of pleadings, namely to inform one's opponent of the case he has to face.

If this knowledge was acquired after the pleading of non-admission was entered, the defendant should simply amend his pleadings on becoming aware of new evidence. The non-admission should either be changed to a denial to or an admission, depending on the new evidence that has come to light. If the pleader fails to amend the pleadings, the pleadings, although technically correct may be untruthful and consequently unethical. Failure to amend one's pleadings should result in the defendant not being permitted to lead evidence unless it is evidence of an expert witness. If the defendant's witness is an expert witness plaintiff would in terms of the rule 36(9) deliver notice not less than 15 days before the hearing of his intention to call the expert witness and not less than 10 days before the trial must have delivered a summary of such expert's opinions and the reasons therefore. Therefore, in terms

of the Uniform Rules of Court, the plaintiff would be informed of the case he faces, albeit by short notice, thus avoiding the plaintiff being taken by surprise.

5 Conclusion

In order to satisfy the requirement of rule 22(2) of the Uniform Rules of Court, that a pleader "shall clearly and concisely state all material facts upon which he relies", the pleader of a non-admission should state why he cannot admit an allegation. This is usually that he has no knowledge. If it appears *prima facie* that the pleader should have knowledge, the pleader must explain why he has no knowledge if a failure to do so would result in his opponent not knowing what case he faces. This serves to place one's opposition in a position to apprise him or herself fully of the defendant's defence, thus fulfilling the purpose of pleadings.

The pleader of a non-admission should be permitted to cross-examine his opponent and his opponent's witnesses with regard to the matters pertaining to the non-admission. The pleader of a non-admission should also be permitted to lead evidence in rebuttal of the matters pertaining to the non-admission if such evidence is that of an expert witness. However, the pleader of a non-admission should not be permitted to lead evidence in rebuttal if that evidence is not the evidence of an expert witness. In short the pleader of a non-admission is only entitled to test the veracity of the plaintiff's case but may not lead evidence relating to the facts or happenings to traverse or contradict the plaintiff's case.

We are of the view that case law supports our conclusions. In *Ntshokomo v Pedi Stores* (1942 EDL 289 298) it was held that the pleader of a non-admission:

> [w]ould not be entitled, either by way of cross-examination or by himself leading evidence, to traverse the accuracy of any of the items in the account for the above period to the correctness of which the plaintiff has deposed.

We agree with this statement save for the bar on cross-examination to test the veracity of the opponent's evidence. Our view is supported in *Standard Bank Factors Ltd v Furncor Agencies (Pty) Ltd* (1985 (3) SA 410 (C) 417-418) where Marais AJ states *obiter* that although he agrees that a pleader of a non-admission may not lead evidence to traverse his opponent's evidence, that it may be going too far to disallow cross examination of one's opponent to test the veracity of his evidence. However, Van den Heever J in *N Goodwin Design (Pty) Ltd v Moscak* (163) disagrees with this view. Her view is that the only distinction between a denial and a non-admission is a matter of emphasis and therefore the pleader of a non-admission is entitled not only to cross examine his opponent but also to traverse the veracity of his evidence with reference to facts by leading contradictory evidence. However Van den Heever J refrains from deciding the issue when she states *obite*r after referring to *Ntshokomo v Pedi Stores* and *Standard Bank Factors Ltd v Furncor Agencies (Pty) Ltd*: "In any event, if it is I who am wrong, both those cases differ

toto caelo from the present…" Clearly Van den Heever J is wrong. The distinction between a non-admission and a denial and the consequences thereof are manifest not only in case law but is also clearly provided for in rule 22(2) of the Uniform Rules of Court. Our contentions are supported by various authors (Becks *Theory and Principles of Pleading in Civil Actions* (2002) 81-82; Morris *Techniques in Litigation* (2010) 96-98; Marnewick *Litigation Skills for South African Lawyers* (2007) 126, 128; Van Blerk *Legal Drafting* (1998) 23).

It is a matter of simple logic that a pleader of a non-admission because of no knowledge should not be allowed to traverse the veracity of his opponent's evidence by leading new evidence. If he has no knowledge how can he lead evidence? If such a pleader were allowed to lead evidence this would be tantamount to encouraging unethical or dishonest pleading, which unethical tendency was noted in *Ngupe v MEC Department of Health and Welfare Eastern Cape* (2006 JOL 16933 SE). It is also a matter of common sense that the pleader of a non-admission should be allowed to cross-examine his opponent.

M S VETTORI
University of South Africa

R J DE BEER
Group 21 Advocates Chambers, Sandton

Onlangse regspraak/Recent case law

Makulu Plastics & Packaging CC v Born Free Investments 128 (Pty) Ltd and Others
2013 1 SA 377 (GSJ)

Interference with contractual relationship – extension of existing categories for purposes of granting relief – requirements for granting of interdict to prevent interference – obligations of landlord in terms of contract of lease as alternative basis for relief

1 Introduction

The delict of interference with a contractual relationship as a species of *damnum iniuria datum*, warranting the institution of an Aquilian claim to enable a party suffering harm in consequence of such interference, has received but scant attention in South African law in recent times. For this reason alone the handing down of a judgment on the basis of this specific form of delict is noteworthy.

The rules of English law pertaining to the tort of "interference with contractual relations" or "inducing breach of contract" have exerted a strong influence on the rules of the South African law of delict in this respect (see eg Van der Merwe *Die Beskerming van Vorderingsregte uit Kontrak teen Aantasting deur Derdes* (1959) 119 sqq, demonstrating a reliance on English precedent in early South African case law; on this action in general, see McKerron *The Law of Delict* (1971) 270-275; Van der Merwe & Olivier *Die Onregmatige Daad in die Suid-Afrikaanse Reg* (1989) 370-382; Neethling & Potgieter *Neethling-Potgieter-Visser Law of Delict* (2010) 306-309; Loubser, Midgley *et al The Law of Delict in South Africa* (2010) 230-233). However, in *Lanco Engineering CC v Aris Box Manufacturers (Pty) Ltd* (1993 4 SA 378 (D) 380D) it was pointed out that "the later cases have sought to bring the claims more strictly under the *actio legis Aquiliae*, broadening its ambit where necessary ...".

In this note an evaluation of the judgement of a full bench of the Gauteng South High Court (Johannesburg) will be given, with specific reference to its value in respect of the development of the delictual claim for redressing harm flowing from an interference with a contractual relationship.

2 Facts and Judgment

The appellant (applicant in the court *a quo*), a tenant of the premises of the first respondent, had, pending the finalisation of an action to be instituted by it, urgently – but unsuccessfully – applied for orders designed to prevent the appellant from: (i) entering into an agreement

with the third respondent (the Ekurhuleni Metropolitan Municipality) in terms of which the third respondent would provide various municipal services to the leased premises; or (ii) requesting, instructing or encouraging the third respondent to terminate the supply of any such services; and (iii) hindering or obstructing the appellant or its employees and invitees in respect of access to and use or enjoyment of the leased property (378J-379B).

The first respondent was the owner of property originally leased to the second respondent who was subsequently placed in liquidation (and although formally a party to the proceedings, played no role in the present application, due to its liquidation). Prior to such liquidation, the appellant was able to render *prima facie* proof of the fact that it had, however, taken occupation of the premises and commenced trading from it with the owner's (appellant's) consent and had even entered into a contract of lease with the owner in terms of an agreement which it had drafted (379E-F, 384D), but remained unsigned. The court held that the signing of the lease agreement remained a mere formality and that it would merely reflect the terms of the oral contract entered into between the appellant and the first respondent (379G). More than a year later the first respondent formally requested the municipality in question (the third respondent) to terminate the electricity supply to the premises occupied by the appellant, stating as reason that its tenant (referring to the previous tenant, the second respondent) had been liquidated (380B). In addition to failing to mention the fact that the appellant was occupying the premises, the first respondent threatened to take legal steps against the municipality and even to claim damages from it if it were to "connect the electricity to any illegal applicant after today ...", prompting the court to conclude: "It is apparent that the first respondent took steps to cause the third respondent to cease supplying electricity to the premises" (380C-D).

As the appellant throughout its occupation of the premises maintained sufficient payments to the municipality (on behalf of the second respondent who had previously entered into a service agreement with the municipality in its capacity as tenant of the premises) to ensure the municipality's continuation of a supply of services, the court concluded as follows (380F):

> The reason why the third respondent terminated the services was not on account of non-payment but on account of the fact that the person with whom it had contracted (the second respondent) had been liquidated. It requested a new user agreement to be concluded and if such agreement were concluded it would continue to supply electricity.

It transpired that the first respondent's attitude in this regard depended on its opinion that the only tenancy that the appellant had had of the premises was on a monthly basis and that such lease had expired pursuant to notice given by the first respondent (381I). Notwithstanding the court's recognition of the dispute between the parties as to the existence or not of a longer lease between the appellant and the first

respondent, it ultimately favoured – for purposes of granting of a temporary interdict – the appellant in deciding that it had provided *prima facie* proof of an existing contract of lease, as mentioned earlier (381B, 384D).

The court's findings in respect of the *prima facie* existence of a lease between the appellant and the first respondent lay the foundation of its decision as to whether the first respondent's behaviour regarding its failure to communicate anything about the existence of a lease between itself and the appellant constituted an actionable wrong. Lamont J (with whom Tsoka J and Francis J concurred), came to the following conclusion (381G-I):

> The acts of the first respondent in notifying the third respondent [the municipality] of the fact that the property was occupied by a person with whom it had no contractual relationship, if the contractual relationship existed, would constitute an interference by the first respondent in the contractual relationship between the appellant and the third respondent. The fact that the contractual relationship had not been concluded, in my view, does not affect the position. It inevitably would have been concluded but for the interference. In terms of the lease agreement the first respondent was by necessary implication, at the very least, to have co-operated with the appellant when the appellant sought to conclude the services agreement with the municipality. It is apparent that the third respondent, in consequence of the interference by the first respondent, declined to conclude a contract with the appellant.
>
> ... In my view, the conduct of the first respondent in performing acts designed to frustrate the free commercial activity of the appellant constitutes a wrongful act.

After referring to case law relevant to the delict of interfering with contractual relationships (382A-384D), the court tested the third respondent's conduct against the established requirements for the issue of an interim interdict (384A-C, which requirements were established in *Setlogelo v Setlogelo* 1914 AD 221, again confirmed in *Hix Networking Technologies v System Publishers (Pty) Ltd* 1997 1 SA 391 (A) and neatly paraphrased by Harms "Interdict" *LAWSA* vol 11 (2nd ed) 419-424) and found it to conform to such requirements (384D-F), indicating that: (a) a clear right on the part of the applicant (appellant) had existed (rights in terms of a lease); (b) the applicant would suffer irreparable harm if the order were not issued (if services were withheld, its occupation of the premises would not be possible); (c) no other form of relief was available; and (d) the balance of convenience favoured continued occupation by the applicant (who had been paying the rent regularly, thereby not causing any harm to the municipality, instead of denying the applicant occupation, involving obviously harmful consequences to the applicant). The court concluded its *ratio decidendi* by emphasising that no risk attached to the position of the first respondent if the municipality were to conclude a services contract with the appellant, because its bylaws made specific provision that the person concluding such contract with

the municipality, and that person alone, shall be liable for any payments in terms of it (384H).

The court finally handed down judgment in favour of the appellant and interdicted the first respondent to prevent the appellant from entering into a services agreement with the municipality, or to request or encourage the municipality to terminate such agreement with the appellant, or to hinder or obstruct the applicant to have access to and enjoyment of the premises. The municipality, as third respondent, was ordered to enter into a services contract with the appellant and to maintain its supply of services in terms thereof. The entire order was made subject to the appellant's instituting an action to ascertain the existence or not of the contract of lease on which it had based its present application, within a month of the date of the present court order (385A-H).

3 Critical Evaluation

3 1 Introduction

In view of the basis on which the court decided to grant the appellant an order as applied for in this case, *viz* the first respondent's infringement of its (potential) contractual relationship with the municipality (the third respondent), regard should at the outset be had to the different situations under which South African law has in the past granted relief to someone who has suffered harm in consequence of a third party's interference with such harmed party's contractual relationship(s). Neethling and Potgieter (306-307) neatly summarise the following instances that have crystallised in our case law:

(i) Where the defendant's or respondent's intentional interference caused one of the contracting parties to commit a breach of contract (a classic formulation of which appears in *Atlas Organic Fertilisers (Pty) Ltd v Pikkewyn Ghwano (Pty) Ltd* 1981 2 SA 173 (T) 202), such as enticing an employee to commit a breach of his or her employment contract (*Dun and Bradstreet (Pty) Ltd v SA Merchants Combined Credit Bureau (Cape) (Pty) Ltd* 1968 1 SA 209 (C) 215G-H; see also case law cited by Neethling & Potgieter 306 nn 261-262; see further Van der Merwe 119-126; Van der Merwe & Olivier 371-373; Neethling *Van Heerden-Neethling Unlawful Competition* 246-247).

(ii) Where, due to the defendant's or respondent's conduct, a contracting party fails to obtain the performance to which he or she is entitled in terms of a contract, although no breach of contract occurs, nor did the defendant's or respondent's conduct amount to an inducement to breach of contract. (This instance is clearly set out in *Lanco Engineering CC v Aris Box Manufacturers (Pty) Ltd* 1993 4 SA 378 (D) 383 G-I, 384E; see further case law quoted by Neethling & Potgieter 307 n 263.)

(iii) Where an act of inducement has occurred, causing no actual breach of contract, but lawful termination thereof. In *Lanco Engineering CC v Aris Box Manufacturers (Pty) Ltd* (383B-E) Galgut J referred to the judgment of Van Dijkhorst J in *Atlas Organic Fertilisers (Pty) Ltd v Pikkewyn Ghwano*

(Pty) Ltd (200D-G) in this regard, where he mentioned the example of a businessman who systematically induces his competitor's employees to leave under circumstances where public policy dictates that if such businessman's aim was not to benefit from those employees' services, but merely to cripple or eliminate the business competitor, such action could well be branded as wrongful (see also Neethling *Unlawful Competition* 247-249).

(iv) Finally, where a contracting party's obligations in terms of his or her contract are increased due to the interference by the defendant or respondent. In this instance it would probably be more accurate to describe the delictual act of the defendant or respondent as merely an "interference with a contractual relationship" and not an "inducement to breach of contract" (see Van der Merwe & Olivier 381 n 6, referring to cases such as *Nieuwenhuizen NO v Union and National Insurance Co Ltd* 1962 1 SA 760 (W); *Saitowitz v Provincial Insurance Co Ltd* 1962 3 SA 443 (W); and *Bester v Commercial Union Versekeringsmaatskappy van SA Bpk* 1973 1 SA 769 (A); Neethling 1981 *THRHR* 78 79).

Neethling and Potgieter (307 *sqq*) correctly place emphasis on the state of modern South African law in this context in terms of which only an intentional inducement to breach of contract or interference with a contractual relationship is actionable. This position was established in the leading judgment of *Union Government v Ocean Accident and Guarantee Corporation Ltd* (1956 1 SA 577 (A), the *ratio decidendi* of which has since been strictly observed in South African case law.

An author such as McKerron (270 *sqq*) treats the case law concerning interference with a contractual relationship in the context of competition law ("interference with trade, business or employment") and relies so heavily on English precedent, that it is virtually impossible to draw a distinction between the rules of English and South African law in this regard. However, as previously mentioned, this part of our law of delict has certainly been heavily influenced by English case law. Although McKerron (272) does not specifically refer to the wrongfulness element in his exposition of the rules pertaining to this delict, he does mention an approach followed in English law that would appear to be pertinent to the establishment of wrongfulness, even in a South African context, namely that courts should not be concerned with the propriety or ethical or social grounds of what parties have done, but should be concerned only with whether the defendant or respondent acted wrongfully (referring to the leading English cases of *Allen v Flood* ([1898] AC 1 142 and *Hodges v Webb* ([1920] 2 CH 70 80). Whether the interference by a defendant or respondent with a contractual relation between two other parties is wrongful, or not, has been held to be determined by means of the general *boni mores* or public policy test (see *Lanco Engineering CC v Aris Box Manufacturers (Pty) Ltd* 380G-H, relying on a range of cases dating from the leading case of *Minister van Polisie v Ewels* 1975 3 SA 590 (A)).

3 2 The Court's Interpretation of the Facts and Application of the Law

3 2 1 Extension of Categories of Interference with a Contractual Relationship

It is evident that Lamont J viewed the facts of the instant case as falling within the confines of the first of the four categories of cases systematised by Neethling and Potgieter (306-307), earlier alluded to (see 3 1 above), *viz* where a defendant or respondent caused one of the contracting parties to commit a breach of contract. This is evident from his general reference to relevant parts of *Lanco Engineering CC v Aris Box Manufacturers (Pty) Ltd* (*supra*) (382A-383A) and more specific reference to *Atlas Organic Fertilisers (Pty) Ltd v Pikkewyn Ghwano (Pty) Ltd* (*supra*) 9383A-384C). It is further clear that the fact that no such contractual relationship, relating to the supply of services, between the appellant and the municipality existed at the relevant time, did not in the least stand in the way of the court's approach on this basis. As pointed out earlier (see 2 above), the court was of the opinion that such contract would inevitably have been concluded, but for the *interference* by the first respondent (381H).

This finding warrants some comment. In this regard it is important to observe that Lamont J referred approvingly (382A-383A) to a long passage from the *Lanco Engineering* case (380B-381B) as authority, *inter alia*, for the broad requirements for a delictual claim based on interference with a contractual relationship, namely "that there must be (a) an unlawful and (b) a culpable (in the broad sense) (c) interference". First, a reference to "interference" seems out of place here.

In the first place, commencing with the last requirement, an interference implies the existence of a relationship, which in the instant case has not yet materialised. Logically there can thus be no interference at all. It is suggested that the word "conduct" would convey the factual situation more accurately. Furthermore, the fact cannot be ignored that, at the time of the conduct of the first respondent, no contractual relation existed between the appellant and the municipality. The bland statement by the court that such a state of affairs does not affect the position, fails to address the legal basis on which such view rests. It is suggested that the mere fact that the first respondent's omission to communicate the "true state of affairs" (according to the appellant, but not necessarily according to the first respondent) to the municipality, should, in itself, be irrelevant in this regard. The legal position regarding omissions is well known in our law: a mere omission attracts no legal consequences, whereas an omission in breach of a legal duty is the only type of omission bringing about legal consequences. The court should at least have explained its finding on the basis of the existence of a legal duty to communicate.

The position taken by the court, in which it regards a potential (future) contractual relationship as having the same standing as an existing legal relationship, resembles the approach of our law of contract regarding the so-called fictitious fulfilment of a condition, *viz* where the law regards a condition as having been fulfilled if one of the parties involved deliberately frustrates the fulfilment of such condition in order to gain a benefit from such negative conduct. However, no such construct exists in our law of delict. Creating further confusion in this regard, is the court's obvious uncertainty with its own finding, reflected in phrases such as " ... if the contractual relationship existed, [it] would constitute an interference..." (381G) and "... assuming the existence of the lease, the appellant would be entitled to obtain relief ..." (381J). The only conclusion to which one must come in evaluating this approach, is that the court in fact extended the categories of instances where the law will grant redress in consequence of an interference with contractual relations, namely where a *potential* or *future* contractual relationship would have come into being, but for the conduct of the defendant or respondent. Should such a notion be tenable at all, it would in any event have to be linked to a further requirement, namely that the defendant's or respondent's conduct in question should have breached a legal duty either to have refrained from acting in such a way that the potential contract did not come into being, or to have taken active steps to promote the conclusion of such a contract.

It is suggested that neither our common law, nor our case law provides any authority for the extension effected in this judgment by the court to the existing law on interference with contractual relationships. The only analogous extension in modern times, relevant for South Africa in view of the history of this type of liability – an extension not even alluded to by counsel or the court – actually took place in England, in the judgment handed down by Lord Denning MR in the case of *Torquay Hotel v Cousins* [1969] 1 All ER 522 (CA) 529 *sqq* where that famous jurist extended the action for inducing breach of contract to a situation where a trade union targeted a hotel by means of industrial action, in the process inducing a supplier who had been contractually bound to deliver furnace oil to it, *as well as a supplier who had not yet entered into a contract for the supply of furnace oil*, to desist from dealing with the hotel in question. However, the general principle enunciated by Lord Denning was criticised as erroneous in the later judgment of *Cheall v APEX* [1983] 2 AC 180. In the case of *Middlebrook Mushrooms Ltd v TGWU* [1993] 1CR 612 620 Neill LJ opined that the liability for inducing breach of contract must not be expanded outside its proper limits, which point of view was finally reiterated in the recent judgment of Lord Nicholls in *OBG Ltd v Allan* [2008] 1 AC 1 [185], in which judgment Lord Denning's view "is exploded", to use the words of Rogers *Winfield and Jolowicz on Tort* (2010) 865. In that judgment (n 32) the court held that one "cannot be liable for inducing a breach unless there has been a breach. No secondary liability without primary liability" (see also Rogers 864). It would thus seem that the present English position is that this form of tortuous

liability is even more restricted than its South African counterpart. Dugdale, Jones *et al Clerk and Lindsell on Torts* (2006) 1496, with reference to *F v Wirral MBC* [1991] Fam 69 115, remark that there are at present four "necessary ingredients" for the tort of inducing breach of contract, *viz* existence of the legal right of a contracting party *vis-à-vis* the other contracting party, knowledge on the part of the tortfeasor of such right and an intention to interfere with it, direct and unjustifiable interference with the right and, finally, damage. This brief comparative review of English sources makes it abundantly clear that the court's approach cannot even be justified by referring to that legal system.

As will be demonstrated later (3 3 below), the only sound legal basis for the finding in favour of the appellant would be an infringement of the rights flowing from the contract of lease concluded between itself and the first respondent.

3 2 2 The Wrongfulness Element

However, even if one were, for argument's sake, to accept the extended basis for granting redress in the case at hand, it is to be doubted whether the appellant should have been successful at all, in light of the two other requirements posed in the *Lanco Engineering* case. In the first instance, one may question whether the element of wrongfulness had in fact been established on the part of the first respondent. From the outset it is quite clear that no illegal means – *viz* the commission of a criminal act or civil wrong – had been employed by it in failing to communicate the position of the appellant to the municipality (*cf* the *Atlas Organic Fertilizers* case 179H, referring to McKerron 270). In view of the general principles pertaining to wrongfulness, this element will in the case of positive conduct only be present where the defendant or respondent infringed the plaintiff's interest in a legally reprehensible way or, in the event of an omission, failed in terms of a legal duty owed to the plaintiff, to prevent the plaintiff from suffering harm (see in general Neethling & Potgieter 33-49; Van der Walt & Midgley *Principles of Delict* (2005) 67-84). In each of these cases the *boni mores* criterion is employed to ascertain whether a factual infringement of an interest, or breach of a legal duty is in fact in a legal sense to be regarded as reasonable and therefore lawful, or unreasonable and therefore wrongful. It is suggested that the application of this criterion – which, as has been demonstrated earlier (see 3 1 above) only comes into being once a factual indication of possible wrongfulness exists, otherwise one would have to do with what is termed "wrongfulness in the air". In this regard Lewis JA uttered the following warning in *Premier, Western Cape v Faircape Property Developers (Pty) Ltd* 003 6 SA 13 (SCA) 31-32: "For an act or omission to be actionable, it must constitute an infringement of a legal interest. Just as there cannot be negligence in the air, so too there cannot be wrongfulness ... in the air ..."; see further sources quoted by Neethling & potgieter 33 n 6.) The only substantiation offered by Lamont J in this case of his finding of wrongful conduct on the part of the first respondent, appears as follows (381I): In my view, the conduct of the first respondent in performing acts designed

to frustrate the free commercial activity of the appellant constitutes a wrongful act.

It is suggested that this is a prime example of an approach to establishing wrongfulness on the basis of "wrongfulness in the air" and that the implications of such an approach is so far reaching, that it could in fact frustrate normal transactions in a society where free trade is the order of the day. In this respect it is well worth to note the warning sounded by McKerron in the context of competition law, where he refers to the approach taken by English judges that courts should not concern themselves with the propriety on ethical or social grounds of what litigating parties have done, but should only be concerned with the question whether the plaintiff could prove such unlawful act on the defendant's part as would entitle him to relief (270 *sqq*, referring to the leading English cases of *Hodgson v Webb supra* and *Allen v Flood supra*; see 3 1 above).

3 2 3 Excursus on Fault in Cases of Interference with a Contractual Relationship

It would appear that, from an academic point of view at least, some measure of doubt exists as to what the fault requirement in this regard entails. Although the present appeal deals with the application for an interdict, for which fault need not be proved by the applicant, the court's extension of the field of application of the remedy for interference with a contractual relation justifies a brief discussion of the present South African position in this regard.

Obviously influenced by the application of the term "malice" in some of the English cases, McKerron (274; quoted *in extenso* in the *Atlas Organic Fertilisers* case 180A-C) is the only South African author who is not averse to an interpretation of this term as being tantamount to "spite", "ill-will" or "improper motive" which, in addition to *dolus*, has to be present on the part of the defendant or respondent and which would then explain that in the case of interference with business relations a person's *motive* has a part to play. In this context McKerron concludes (274):

> [C]ourts of law should not attempt to distinguish between acts of interference which are unfair and unreasonable. But a line can and should be drawn between acts of interference whose object is the defence or advancement of a person's own interests, and acts of interference whose sole or dominant purpose is the infliction of harm for harm's sake.

It is suggested that if this line of reasoning is followed, the facts of the present case do not warrant a finding of malice on the first respondent's part. The sole fact that a measure of doubt existed as to the existence of a contract of lease between the appellant and the first respondent (that it "is apparent that there is a dispute between the first respondent and the applicant as to the existence of the lease") and that the court at most found that there was *prima facie* evidence of the appellant's occupation

of the premises in terms of a lease (384D), to my mind definitely does not justify a finding of malice on the part of the first respondent. If the court's reference to the first respondent's conduct as "designed to frustrate the free commercial activity of the appellant" (381I, quoted above) was meant as a positive finding as to malice, it is clearly wrong. (For stringent criticism of an interpretation of malice as representing an improper motive, see Van der Merwe 13 *sqq*.)

There seems to be a greater measure of unanimity in respect of the requirement of fault: intent (*dolus*) at least is required, with specific emphasis on the element of consciousness of wrongfulness (see the *Lanco Engineering* case 380J, merely endorsing the earlier leading judgment of *Union Government v Ocean Accident and Guarantee Corporation Ltd* (*supra*); Neethling & Potgieter 308-309 who, in principle, favour negligence as a sufficient form of fault). It is suggested that if the facts of this case are kept in mind, it is doubtful whether any action on the first respondent's part giving rise to the appellant's dilemma had been done with full consciousness of wrongfulness, considering the great measure of uncertainty regarding the existence or a lease between the appellant and the first respondent. It could well be argued that the first respondent's conduct had simply been consistent with an opinion that no lease agreement had been concluded and that it simply acted to its own advantage in wishing to see the appellant leaving the premises.

3 2 4 Compliance with Requirements for Granting of an Interim Interdict

Finally, something falls to be observed on the way in which the court applied the requirements for the granting of an interim interdict to the appellant. Courts confronted with instances of interference with a contractual relationship usually have to decide a claim for damages instituted by the aggrieved party. However, Van der Merwe (195) points out in this respect that "die krag van 'n interdik ook nie uit die oog verloor [moet] word nie". This is indeed the only reported case of this kind which I could find and in which the remedy of an interdict featured. The first, second and fourth requirements regularly posed for the granting of a temporary interdict (see 384A-C; see further *LF Boshoff Investments (Pty) Ltd v Cape Town Municipality* 1969 2 SA 256 (C) 267; Neethling & Potgieter 255; Van der Merwe & Olivier 257) call for some comment.

The first requirement is that the applicant should prove a *prima facie* right (which is being infringed or imminently threatened by the respondent). If the typical position in cases of this kind is taken into account – where A and B are in a contractual situation on which C infringes – it will be evident that the court did not pay careful attention to this. In this case one could imagine A being the appellant (tenant), B being the municipality and C being the first respondent (landlord). In the typical case the *prima facie* right would be that of A, flowing from his contractual relationship with B, under threat by C. The court's positive finding in respect of the existence of the required *prima facie* right,

appears as follows: "The appellant [A] has established prima facie that it occupies the property pursuant to a lease [with landlord C]" (384D). It is obvious from the information supplied in square brackets, that the required *prima facie* right as explained above – *viz* between A and B – is totally lacking! The only right which the court alluded to here, was B's right *vis-à-vis* C in terms of the purported contract of lease. This is a far cry from the required right of A *vis-á-vis* B, on which C's conduct could constitute an infringement. The unavoidable conclusion from this is that there had been no compliance in this case with the first requirement for granting an interim interdict.

As Harms (*LAWSA* vol 11 422) points out, the second requirement for obtaining an interim interdict, *viz* that irreparable harm would follow if such interdict were not granted, need not *eo nomine* be adhered to where the applicant has already established a clear right; and that it is furthermore closely related to the last requirement posed for the granting of an interim interdict, such being a consideration of the balance of convenience, which should favour the granting of the interdict. It is trite that this means that such interdict should be granted if this would not entail serious detriment to the respondent, while a refusal of the order would give rise to the applicants suffering substantial harm (*Olympic Passenger Service (Pty) Ltd v Ramlagan* 1957 2 SA 382 (D) 383F; see further Harms *LAWSA* vol 11 422, in particular authorities quoted in nn 1-4). It should be kept in mind that an interim interdict, when granted, only applies pending the outcome of a trial action based on the same facts (see Neethling & Potgieter 255). From the order granted by the court it appears that it was expressly made subject to such action being instituted within a month of the granting thereof. One could thus safely predict that the dispensation brought about by the granting of the order in favour of the appellant would be of relatively short duration. Was that the main reason for the court's lenient attitude towards the appellant? One can only speculate on this, as the judgment itself sheds no further light on the matter.

Furthermore, it is not clear on what basis the order against the municipality (the third respondent) to conclude a contract for the provision of services with the appellant rests, having regard to the above-mentioned requirements. In the first place, from the facts it is patently clear that a contractual relationship had not yet come into existence between the appellant and the municipality; in fact, the entire object of the legal proceedings had been to effect such a relationship. Any *prima facie* right of the appellant required *vis-à-vis* the municipality on the basis of a services contract had thus been totally absent. Although it might be argued that, in the event of the existence of a lease between the appellant and the first respondent (as *prima facie* decided by the court), the municipality would have been obliged to conclude a contract for the provision of services with the appellant (on the basis of a consumer's right based on the relevant bylaws), there is no indication in the facts set out in the judgment that the municipality would not have concluded such a service agreement if it had been informed by the first respondent of the

appellant's tenancy of the premises in terms of a lease. As a necessary corollary to the existence of a clear or *prima facie* right on the part of an applicant, our law requires an act or omission on the part of the respondent indicating that a breach of the right would occur in the absence of an interdict (see eg *Conde Nast Publications v Jaffe* 1951 1 SA 81 (C); Van der Merwe & Olivie 252). It is therefore suggested that this fundamental requirement for granting the order against the municipality was lacking. One can justifiably argue that the appellant could have proceeded against the municipality with an urgent application for a temporary interdict aimed at the conclusion of a services contract, only after the municipality would have failed to conclude the contract for the provision of services with the appellant after having been informed of the court order against the first respondent.

3 3 An Alternative Basis for Granting of the Interdict

Accepting the correctness of the court's granting the appellant an interim interdict against the first respondent, one is forced to consider the correct basis on which such relief was granted. It is suggested that this basis emerges clearly from the court's own finding in respect of the presence of the first requirement for the granting of an interim interdict, *viz* that "the appellant has established *prima facie* that it occupies the property pursuant to a lease" (384D). In terms of the legal rules governing leases, it is well established that the landlord is under a legal obligation to grant the tenant occupation of the leased premises, together with all accessories that are essential for its proper use and enjoyment (*McNeil v Eaton* (1903) 20 SC 507 512; *Pistorius v Abrahamson* 1904 TS 643 648; see further Du Bois *et al Wille's Principles of South African Law* (2007) 911; Kerr & Lotz "Lease" *LAWSA* vol 14 (First Reissue) 169 sqq). This entails a performance by the landlord of all actions necessary to enable the tenant to occupy the premises in terms of the contract.

It is suggested that such performance of the first respondent *in casu*, on the basis of the *prima facie* contract of lease with the appellant, was to inform the municipality of the lease, to enable it to supply services to enable the appellant (tenant) to occupy the premises as contemplated in the lease contract. The flipside of this obligation is of course the tenant's right to such performance against the landlord. It is suggested that it is in fact this right that formed the *prima facie* right required for the granting of the interdict by the court.

4 Conclusion

As pointed out previously (1 above) this is the first judgment on interference with a contractual relationship to have appeared for a long time and, in addition, it is a judgment of a full bench, which has implications in terms of the doctrine of precedent. This entails that at least the South Gauteng High Court (Johannesburg) will in future be bound to recognise a considerable extension to the instances where certain conduct is to be considered to constitute such an interference, *viz*

that a mere potential (future) contractual relationship will have the same standing as an existing legal relationship.

As pointed out, serious doubt can be expressed on the correctness of the court's reasoning in this regard on both logical and legal grounds. It is suggested that the applicant (appellant) should have been granted relief against the first respondent, but then on the simple basis of the contract of lease that was found to exist between them. It is to my mind understandable that the court *a quo* refused the application on the basis of interference with contractual relations.

It would appear that the case was wrongly argued but that the full bench notwithstanding wished to grant temporary relief to the applicant, which would seem equitable under the circumstances (especially in view of the temporariness thereof, and in light of the importance attached to the balance of convenience). If I am correct in my assessment of this aspect of the case, this is a prime example of the old adage that hard cases make bad law.

J SCOTT
University of Pretoria

Boekaankondigings/Book Announcements

Politokrasie: 'n Peiling van die Dwanglogika van die Territoriale Staat en Gedagtes vir 'n Antwoord Daarop

deur

KOOS MALAN

Pretoria University Law Press 2011 vii en 358 bladsye

Elektronies gratis beskikbaar by http://www.pulp.up.ac.za/pdf/2011_01/2011_01.pdf

Koos Malan is nie iemand wat in sy geskrifte oor die reg en verwante dissiplines bloot 'n beskrywende weergawe gee van die voorwerp van sy ondersoek nie. Ek sou hom ook nie tot die genre beperk van outeurs wat agter die empiriese substansie van dit wat is, na 'n historiese basis, teoretiese begronding en/of eties gefundeerde waardesisteem op soek gaan nie. Koos Malan is 'n diep denker wat fundamenteel gewortelde en kwalifiserende werklikheidsmomente ontbloot en wat die sin en betekenis van instellings, norme, en waardes van die wêreld waarin ons leef, kwalifiseer. "Politokrasie" of "'n politokratiese orde", afgelei van die klassieke Griekse begrippe *polis* (staat) en *politai* (burgers) (276), gaan op soek na "'n politiek-juridiese orde" (10), en meer bepaald na dit wat bepalend is vir die dieper-liggende realiteit van daardie orde. Inhoudelik gaan dit om die territoriale staat en ons verbondenheid aan 'n politieke orde – histories en kontemporêr.

Politokrasie bevat 'n rykdom van historiese gegewens rondom die vestiging en ontwikkeling van 'n territoriale staat wat veranker is in 'n bepaalde godsdienstige vooroordeel en daaropvolgende sekularisasie van die gangbare politieke denke, maar uiteindelik beslag begin kry het met die Vredesverdrag van Augsburg wat die Dertigjarige Oorlog (1618-1648) in Europa tot 'n einde gebring het. Ontwikkeling van die staatsbegrip het natuurlik nie daar opgehou nie, en binne die paradigma van politokrasie, ontleed die skrywer na-middeleeuse en kontemporêre waardesisteme wat die *res publicae* kwalifiseer. *Res publicae* in dié sin word nie deur die skrywer tot die sakeregtelike betekenis van daardie begrip beperk nie, maar beteken vir hom meer bepaald "die gemenebes – openbare sake van algemene belang vir alle burgers" (291).

Kortom, die leser wat miskien nie sin het aan die diep denke van Koos Malan nie, sal nietemin veel baat kan vind by die rykdom aan historiese gegewens wat in *Politokrasie* vervat is. Daarmee gaan saam besonder waardevolle verwysings na die denke van rigtinggewende politieke filosowe, waaronder Thomas Hobbes, John Locke, Hugo de Groot, Jean-Jacques Rousseau, John Rawls en vele andere. *Politokrasie* bied die leser

ook waardevolle inligting oor basiese staats- en regsfilosofiese waardesisteme, waaronder die instelling en toepassing van staatlike demokrasie, die leerstuk van menseregte en die reg op selfbeskikking van etniese, geloofs- en taalgemeenskappe.

Burgerskap is ook veel meer as net 'n bepaalde openbare identiteit en nasionale groepsnaam, of die aanspraak op 'n paspoort. Dit behels "mede-regering van die *res publicae*", "die daadwerklike en deurlopende deelname aan, en verantwoordelikheid neem vir die *res publicae*", "die bekleding van 'n amp, waarvolgens burgers deur aktiewe handeling daadwerklik vir die *res publicae* verantwoordelikheid aanvaar en daaroor beskik" (279). Miskien sou die skrywer ook nog ietwat dieper kon delf om 'n bepaalde psigiese samehorigheidsgevoel van nasieskap te ontbloot; miskien dit wat deur Von Savigny as die volksgees, 'n grondliggende nasionale ethos, geïdentifiseer is en wat sy stempel afdruk op ons nasionale bewussyn en regsinstellings en besielend in die tydsgees van 'n bepaalde era veranker lê. Professor Silvio Ferrari van die Universiteit van Milan verwys daarna as 'n burgerlike religie ("civil religion"); nie net die "koue verwisseling van regte en verpligtinge" nie, maar "iets meer, wat die harte van die burgers verwarm" (Ferrari "Introduction: Civil religions: Models and perspectives" 2010 *Geo Wash Int LR* 749). En daar is ook 'n belangrike les te leer uit die op skrif gestelde herinneringe van Professor Karl Doehring, wat jare lank aan die Max Planck Instituut in Heidelberg, Duitsland verbonde was en wat hy kort voor sy dood voltooi het: Geleerdes van die huidige geslag kan wel akademies besin oor beleidstandpunte en gebeure van die verlede; maar as hulle dit nie self beleef het nie, kan hulle nie die begronding van die leerstellinge, gewortel in 'n bepaalde tydsgees, volkome na waarde skat nie (Doehring *Von der Weimarer Republik zur Europäischen Union Erinnerungen* (2008)).

Miskien dan nog net enkele punte van kritiek.

Wat die volkeregtelike vereistes vir staatskap betref, verwys die skrywer slegs na die Montevideo beginsels, oftewel die verklarende teorie van staatskap (100). Daar is egter ook die konstituerene teorie wat die klem laat val op erkenning van 'n gebied se staatskap deur ander lede van die internasionale gemeenskap van state. John Dugard het die konstituerene teorie van die ander kant bekyk deur daarop te wys dat nie-erkenning deur die Verenigde Nasies onomwonde bewys is dat 'n gebied wat op staatskap aanspraak maak, nie vir staatskap vir volkeregtelike doeleides kwalifiseer nie (Dugard *Recognition and the United Nations* (1987)). Ek het self aan die hand gedoen dat daar 'n onderskeid getref moet word tussen staatskap in inter-individuele verhoudings tussen state aan die een kant en vir die doeleindes van staatlike gemeenskapsverhoudinge aan die ander kant. 'n Gebied wat ongeregverdig op staatskap aanspraak maak (bv die Turkse Republiek van Noordelike Ciprus) kan in onderlinge diplomatieke, handels- en ander internasionale betrekkinge tree met 'n staat wat daardie gebied se staatskap erken (Turkye), maar sal nie deel uitmaak van die internasionale gemeenskap van state wat lede kan word van die

Verenigde Nasies, of wie se doen en late in berekening gebring sal word vir die skepping van volkeregtelike gewoontereg nie (Van der Vyver "Statehood in international law" 1991 *Emory Int LR* 9).

Wat woordkeuse betref, is daar oor die algemeen nie fout te vind met die onderskeid wat die skrywer maak tussen volk ('n etniese of kulturele gemeenskap) en nasie (gebaseer op burgerskap van 'n bepaalde staat) nie. Waar hy egter sy eie voorkeur vir die begrip "territoriale staat" eerder as "nasiestaat" regverdig, motiveer hy dit op grond daarvan dat "'n soewereine politieke organisasie binne 'n vasgestelde gebied" gevestig kan word "ongeag of daar enige kulturele, linguistieke, godsdienstige of etniese gemeenskaplikheid by die staatsbevolking aanwesig is" (9). Hoewel dit waarskynlik nie so bedoel is nie, kan die argument geïnterpreteer word as sou "nasie" met "kulturele, linguistieke, godsdienstige of etniese gemeenskaplikheid" geïdentifiseer moet of kan word.

Etniese, godsdienstige en taalgemeenskappe is inderdaad as subjekte van die volkereg van groot belang, naamlik as die draers van 'n reg op selbeskikking (vgl ook arts 31 en 185-187 van die Grondwet van die Republiek van Suid-Afrika), en in *Politokrasie* word pertinent daarna verwys met 'n oorsigtelike analise van die geskiedenis en betekenis van daardie konsep in die volkereg. Die skrywer het egter miskien nie volkome rekenskap gegee van die verskillende betekenisse van "self-beskikking" soos wat dit sedert die Eerste Wêreldoorlog beslag gekry het nie – miskien omdat hy hom te beknop van 'n publikasie van Cherif Bassiouni oor die onderwerp laat bedien het.

Die term "selfbeskikking" het eintlik sy oorsprong in kommunistiese literatuur. Volgens Karl Marx en Friedrich Engels sou die ganse wêreld mettertyd deur bemiddeling van 'n revolusie van die proletariaat aan 'n ekonomiese struktuur onderwerp word waarin private eiendomsreg verbode sou wees en alles aan almal sou behoort. Die vraag het egter ontstaan oor wat die status sal wees van nasionale state binne hierdie, ononderhandelbare, oorkoepelende ekonomiese struktuur. In afsonderlike geskrifte van Joseph Stalin (*Marxism and the national question* (1913)) en Lenin (*Thesis on the socialist revolution and the right of nations to self-determination* (1916)) word aan die hand gedoen dat hulle 'n reg op selfbeskikking sal hê. Die relevansie van die begrip in die volkereg word egter gangbaar aan die Amerikaanse President, Woodrow Wilson, toegeskryf. In sy Veertien Punte Toespraak van 8 Januarie 1918 in die Kongres (die Amerikaanse wetgewer) verwys hy onder meer na

> *[a] free, open-minded, and absolutely impartial adjustment of all colonial claims, based upon a strict observance of the principle that in determining all such questions of sovereignty the interest of the population concerned must have equal weight with the equitable claims of government whose title is to be determined.*

Hoewel hy in daardie toespraak nooit die woord "selfbeskikking" gebruik het nie, het hy later wel op 11 Februarie 1918 in 'n toespraak voor 'n gesamentlike sitting van die twee Kamers van die Kongres verklaar:

> *National aspirations must be accepted; people may now be dominated and governed only by their own consent. 'Self-determination' is not a mere phrase. It is an imperative principle of action, which statesmen will henceforth ignore at their peril.*

Die reg op selfbeskikking het daarop praktiese beslag gekry in die mandaatstelsel van die Volkebond waaronder die toekomstige status van nasionale state wat deel was van die wêreldryke wat in die Eerste Wêreldoorlog verslaan is, beslag moes kry. Die mandaatstelsel is terloops deur Generaal JC Smuts van Suid-Afrika (hy was deel van die Britse afvaardiging by die vredesonderhandelinge in Versailles) in 'n geskrif van Desember 1918 onder die opskrif "A practical suggestion" ontwerp (Smuts was ook die outeur van die voorrede tot die Handves van die Vernenigde Nasies). Die 1918 geskrif is deur President Woodrow Wilson by sy tweede ontwerp van 'n Statuut vir die Volkebond ingesluit en die mandaatstelsel het toe deel geword van daardie Statuut. Die reg op selfbeskikking (meestal politieke onafhanklikheid) van die nasionale state waarom dit gegaan het, moes onder die leiding en toesig van 'n mandaadstaat met verloop van tyd verwerklik word. Jan Smuts, soos almal weet, was egter ook maar 'n ou jakkals. Sy ontwerp het vir net een C mandaatgebied voorsiening gemaak (Suid-Wes Afrika), wat daarin van die A en B mandate verskil het dat dit bestem sou wees om eventueel by die mandaatstaat (die Unie van Suid-Afrika) ingelyf te word as 'n vyfde provinsie.

Na die Tweede Wêreldoorlog het die klem van die selfbeskikkingsparadigma geskuif na die reg van koloniale gebiede op politieke onafhanklikheid. In die sestiger jare is die begrip ook aangewend om die reg te kenteken van bevolkingsgroepe wat onderworpe was aan 'n rassistiese bewind om deel te hê aan die regering- en wetgewingstrukture van die land waartoe hulle behoort. Hierdie variant van die begrip het voortgespruit uit die standpunt van die Suid-Afrikaanse regering dat die tuislandbeleid in ooreenstemming is met die reg op selbeskikking van die etniese groepe wat op onafhanklikheid aanspraak gemaak het. Die Verenigde Nasies het die standpunt verwerp omdat die tuislandbeleid op die swart bevolking afgedwing is deur 'n regering waarin hulle geen verteenwoordiging gehad het nie.

'n Vierde, en die huidiglik oorheersende, betekenis van selfbeskikking behels die reg van 'n etniese of kultuurgemeenskap om sy eie kultuur te bevorder, van 'n godsdiensgroep om sy godsdiens te beoefen, en van 'n linguistieke gemeenskap om sy taal te praat, sonder ongevraagde en onregverdigbare staatsinmenging. Trouens, die volkereg plaas op die regerings 'n plig om die bevordering van 'n kultuur, die beoefening van 'n godsdiens en die gebruik van 'n taal te ondersteun en te bevorder.

Dit moet beklemtoon word dat die omskrywing van die reg op selbeskikking in Artikel 1 van die *Internasionale Verdrag met betrekking tot Burgerlike en Politieke Regte* en Artikel 1 van die *Internasionale Verdrag met betrekking tot Ekonomiese, Sosiale en Kulturele Reg*te van 1966, wat

die reg van "alle bevolkingsgroepe" insluit "om vryelik hulle politieke status te bepaal" aangeneem is gedurende die tyd toe die klem van die reg op selfbeskikking op dekolonialisme geval het en daarom nie lukraak aangewend kan word om die substansie van die reg op selfbeskikking van etniese, godsdienstige en taalgemeenskappe te bepaal nie. Die internasionale gemeenskap is self nie altyd hierop bedag nie. Die Verenigde Nasies se *Verklaring met betrekking tot die Regte van Inheemse Bevolkingsgroepe* van 2007 gebruik daardie einste woorde om die reg op selfbeskikking van inheemse bevolkingsgroepe te omskryf (art 3), en omdat die reg van 'n bevolkingsgroep om vryelik oor sy politieke status te beslis op politieke onafhanklikheid dui, het vier lande teen aanvaarding van die Deklarasie gestem, naamlik Australië, Kanada, Nieuw Zeeland en die Verenigde State van Amerika. Die Deklarasie vermeld egter self dat geeneen van sy bepalings geïnterpreteer moet word as sou dit geoorloof wees "om die territoriale integriteit of politieke eenheid of soewereiniteit van onafhanklike state te verbrokkel" nie (art 46.1). Om die reg op selfbeskikking van bevolkingsgroepe met politieke onafhanklikheid te vereenselwing maak in elk geval geen sin nie. Dit is per slot van sake 'n territoriale gebied wat in gegewe omstandighede op politieke onafhanklikheid aanspraak kan maak en nie 'n bevolkingsgroep nie.

Ten slotte het ek ietwat kommer omdat die skrywer hom by tye van 'n hoogdrawende skryfstyl laat bedien wat ook anglisismes insluit. Dit gaan nie hier om oorspronklike woordskepping om ongekende begrippe te bekenteken of om meersinnigheid van 'n bestaande woord te vermy nie. Daarvoor is die skrywer goed gekonfyt, soos wat ook uit die titel van sy boek blyk. Ek kom egter uit 'n tyd toe die gebruik van anglisismes as Afrikaner hoogverraad beskou is – behalwe natuurlik as die Suid-Afrikaanse Akademie vir Wetenskap en Kuns besluit het dat dit voortaan ô-raait sal wees as jy hierdie of daardie verafrikaanste nabootsing van 'n Engelse woord of uitdrukking sou gebruik (die Akademie kon "bromponie" vir "scooter" bemark, maar moes tou opgooi met "melkskommel" ipv "bruismelk" of "as volg" ipv "soos volg"). Daar moet egter nog steeds kapsie gemaak word teen die swak Afrikaanse woordgebruik van mense soos Johan Stemmet van Noot-vir-Noot faam (bv deur 'n gevolgtrekking in te lui met "So, ..." ipv "dus" of "daarom"). Maar, om terug te keer tot *Politokrasie*, kan met die volgende paar voorbeelde volstaan word: "die politiek-sosiale kohesie ..." (6); "kan ook na Augustinus teruggespeur word" (15); "oor partikuliere grense" (17) "die Engelse spesie van regsoewereinitiet ..." (25); "die grille en voorkeure van ..." (105); "partikuliere gemeenskappe" (275). Ek sou ook "basiese norm" bo "grondnorm" verkies as 'n vertaling van *Grundnorm* (vgl 102) – en let daarop dat Duitse selfstandige naamwoorde met 'n hoofletter gespel word. Eenvoudige en algemeen verstaanbare Afrikaans bly steeds die beste medium, ook in akademiese geskrifte.

Koos Malan verdien egter 'n klop op die skouer en 'n stywe handdruk vir 'n weldeurdagte en indringende analise van die geskiedenis en essensiële eienskappe van die territoriale staat. Daar is, so lyk dit my,

deesdae 'n algemene tendens om regsopleiding uitsluitlik praktykgerig in te klee ten koste van die historiese, teoretiese, en wysgerige begronding van die reg as wetenskap. Dit is jammer. *Politokrasie* sal die leser daaraan herinner dat universiteite bestem is om akademiese instellings te wees en nie bloot in tegnikons omgeskep moet word nie.

Johan D van der Vyver
Emory University

Africa and the Deep Seabed Regime: Politics and International Law of the Common Heritage of Mankind

by

Edwin Egede

Heidelberg Springer 2011 xxxii plus 271 pages

Price: Euro 114.99

In *Africa and the Deep Seabed Regime: Politics and International Law of the Common Heritage of Mankind,* the author Edwin Egede explores the role of Africa, the second largest of the seven continents on earth, in the development and establishment of the regime of the deep seabed beyond national jurisdiction and the concept of the Common Heritage of Mankind. A salient feature of this book is that the author delves into examining the intriguing and lopsided regime of the deep seabed beyond national jurisdiction and the concept of the Common Heritage of Mankind exclusively from the viewpoint of African states. The value of this particular work lies in making an effort to fill up the vacuum in the existing academic literature as to the role of African states in the development of the regime of the deep seabed beyond national jurisdiction and the concept of the Common Heritage of Mankind. The author examines African States' contributions to the evolution and development of the international law norms related to this regime and places these contributions in the context of vital historical, social, political and economic factors that generally influence African states' attitudes to international law.

The book is stratified into eight chapters and an introductory part. At the outset, the introductory part indentifies the rationale of the topic of the book and later on, arrays the discourse of subsequent chapters in proper context. Chapter 1 dapples with historical contribution of the African states to the development and evolution of the deep seabed regime and the concept of Common Heritage of Mankind. The continent of Africa is delimited on the north by the Mediterranean Sea, the west by the Atlantic Ocean, the northeast by the Red Sea and the southeast by the Indian Ocean. Thence, the sea is of great significance to the continent.

The chapter seeks to examine the historical developments of the law of the sea and deep seabed regime from an African perspective. The newly independent African states, at the outset of 1960s, were grumbled as to the traditional law of the sea which was more favorable to the technologically developed maritime states. Such disparity also revealed the tussle at the international sphere swelling between the rich technologically developed states and the generally poor, less technologically developed African and other third world states.

Chapter 2 analyses the role of African states in arraying the delineation of the deep seabed regime and analyses the geographical significance of that ocean space by the wordings of the Law of the Sea Convention. Since the deep seabed area is distinct from seabed and subsoil with national jurisdiction, it is vital to figure out the scope of the concept of the common heritage of mankind in that deep seabed area. Hence, the African coastal states, which are 39 in number, have a pivotal role to play in delving upon the issue.

Chapter 3 discusses the viewpoint of African states regarding the legal status of the concept of "common heritage of mankind", a relatively novel concept in the realm of international law and international relations. The concept becomes part of customary international law through the various General Assembly resolutions adopted in the 1960s and 1970s. The concept of "common heritage of mankind" is contained in articles 137 and 141 of the Law of the Sea Convention and imposes a binding duty upon all states to use the deep seabed area only for peaceful purposes and to refrain from claiming or exercising sovereignty over the deep seabed area or its resources. The concept has been inducted as a human right in article 22(1) of the African Charter on Human and Peoples' Rights. African States, along with other developing States in the Group of 77, take the view that the concept has the character of *jus cogens* (70).

Although African states are more prone to illustrate the concept of common heritage of mankind as incorporating the idea of common property, there are a motley of interpretations of the concept depending upon ideological, political and economic interests of the various states around the globe. Therefore, the author refers to "common heritage of mankind as a concept ambiguous in nature (60). The concept precludes States from appropriating the region further and excludes unilateral exploitation. Instead, it requires any exploitation of the area should be collective and regulated.

Chapter 4 examines the regime of the deep seabed area under Part XI of Law of the Sea Convention, as modified by the Agreement of 1994, an amending instrument that paves free market economy, with the aim of pinpointing certain effects on African states. Rather than exploring the modifications in the institutions and framework system of mining, the chapter analyses the reasons behind African states' preparedness to go along with the 1994 Agreement, considering their original and

unequivocal position during the UNCLOS III on the issues changed by the Agreement. The Agreement has significantly altered the original Part XI provisions of the Law of the Sea Convention by modifying its provisions on transfer of technology, production policy, financial terms of contracts and the review conference. Such modifications evidently favor the industrialised developed states to the disadvantage of African states and other developing states. In this chapter, the author analyses the effect of such changes on African states.

As a fitting sequel, Chapter 5 delves in examining the institutional framework of the regime established by Part XI of the Law of the Sea Convention, as modified by the 1994 Agreement. The chapter is bifurcated into two parts. The first part examines the institutional framework of the regime and the role of African states. This part is also divided into two sub-parts, Part A deals with the International Seabed Authority (ISA), while Part B deals with the dispute settlement mechanism. The second part overviews the issue of funding of these institutions and the contributions of African states. To inquire as to the ambit of representation of African states at the organs of institutions like International Seabed Authority, the author scours to discover whether the institution reflects effective democratic participation by all States. His findings reiterate that although well represented in the realm of the institutional framework of the Deep Seabed Area, African states do not have an effective voice matching with their role and significance.

Chapter 6 overviews the system of mining of the common heritage of mankind resources in the area by looking at the system of exploration and exploitation of the seabed with an aim of analysing contributions of African states to this and some possible effects on such states. The author evaluates the potential of African states' participation, along with its landlocked and disadvantaged states, on south-south co-operation with other developing states in deep seabed mining activities.

The financial and technological requirements burdened by the Law of the Sea Convention in tandem with the 1994 Implementation Agreement and the Mining Code wring the potential of African states with limitations. In Chapter 7, the author explores various possible co-operative actions that African states may embark upon to overcome the bulwarks in order to advance the prospects of their actual, direct and effective participation in deep seabed mining activities. Although the crisis of non-participation can easily be diluted by other pressing social problems that are currently entangling the African continent, such as poverty alleviation, the fight against HIV/AIDS and the provision of basic amenities, the author opines that the non-participation of African states in deep seabed mining activities is more due to a lack of interest on the part of African states rather than those daunting and entwining constraints.

The concluding chapter 8 is stratified into two parts. In the first part, the author outlines his appraisal of African states' contribution to the

common heritage of mankind. The comprehensive study of the arena of the deep seabed area regime reiterates the chieftain contribution of African states. The author identifies the impetus of African states in the vaunted regime of UNCLOS III as a "stamp of influence" as well as a "power-show" against the backdrop of western states' domination in the orb of law of the sea. To rev up African states' participation in deep seabed activities, the author propounds a brisk list of recommendations that refers to forming an articulated ocean and coastal policy; an institutional framework; and to take measures for effective participation of African states, as per article 148 of Law of the Sea Convention, in deep seabed mining activities.

In the second part of the concluding chapter the author indicates that beneath the froth of legal and political issues the tussle between power and justice is continuing in the coliseum of international law and politics. The African and developing states advocated for a justice-based approach, through which weak and deprived states can advance claims in the face of power, to protect the deep seabed area and its resources from approach of exploiting as much of the resources of that area, which the industrialised developed states contended for under the rubric of a "first come, first served basis". Although a justice based approach is upheld in Law of the Sea Convention, it is the 1994 Agreement that dilutes the impact of the common heritage of mankind concept by limiting the impact of the justice based approach as far as the behemoth states are willing to accept.

This book is an accomplishment which puts forward a distinctive African flavour to the discourse of the regime of the deep seabed beyond national jurisdiction and the concept of the common heritage of mankind. The merit of this book under review lies in the fact that it sheds light on making an emphasis that the impetus of Africa paves an avenue to attempt to resolve outstanding north-south issues related to economic and social development. In this respect at least, the book does represent something of a novelty in legal literature.

Mohammad Rubaiyat Rahman
South Asian University

Redaksionele Beleid

1. *De Jure* se hoofoogmerke is die bevordering van die regswetenskap en 'n kritiese analitiese benadering tot die reg. Ten einde hierdie doelwitte te verwesenlik, word oorspronklike bydraes van 'n hoë akademiese standard daarin gepubliseer.

2. *De Jure* is 'n nasionale en algemene regstydskrif, waartoe akademici, lede van die regsbank en lede van die verskillende regsprofessies bydra. Geen voorkeur word aan outeurs van enige instelling verleen nie. Die besluit om 'n voorlegging te publiseer, hang daarvan af of dit aan die hoë kwaliteitsvereistes van *De Jure* en of daar plek vir die publikasie beskikbaar is.

3. Voorleggings in beide Afrikaans en Engels word gepubliseer en dit kan uit artikels, aantekeninge en resente regspraak en boekresensies bestaan. 'n Vertaalde title en 'n kort opsomming van ongeveer 300 woorde in Engels moet 'n artikel vergesel indien die artikel in Afrikaans geskryf is. In geval van 'n artikel in Engels, word dieselfde vereiste gestel met betrekking tot 'n Afrikaanse title en opsomming.

4. Ten einde vir publikasie oorweeg te word, moet 'n voorlegging die resultaat van oorspronklike navorsing deur die outeur(s) wees, aan die toepaslike regsbeginsels met betrekking tot die publikasie voldoen (soos outeursreg ensovoorts), iets wat voldoende nuut is tot die bestaande literatuur voeg en aan taalkundige, tegniese en styl-vereistes van publikasies in *De Jure* voldoen. Outeurs is persoonlik daarvoor verantwoordelik om toe te sien dat hul voorleggings aan al hierdie vereistes voldoen.

5. Voorleggings word slegs vir oorweging aanvaar op die basis dat alhoewel die redaksie die finale besluit oor publikasie maak, voorleggings aan gepaste portuur-evalueting en evaluering deur deskundiges onderwerp word, asook aan evaluering deur lede van die advieskomitee indien nodig. Die redaksie behou die reg voor om alle voorleggings wat ivr publikasie aanvaar is, ooreenkomstig die redksionele beleid te redigeer, asook om dit te verkort indien nodig.

6. Manuskripte mag slegs in elektroniese format (gebruik MS Word) via 'n e-pos aan die redkteur of deur aflewering van 'n elektroniese skyf by die kantoor van die redakteur, voorgelê word. Bykomend, moet outeurs

(a) relevante kontakbesonderhede, in besonder e-posadres(se) en telefoonnommers verskaf;

(b) waarborg dat hulle geregtig is om die volle voorlegging te publiseer en dat dit, of deel daarvan die elders gepubliseer is nie;

(c) openbaar of die voorlegging, of deel daarvan, ook aan 'n ander tydskrif vir publikasie voorgelê is; en

(d) onderneem om redelike kennis aan die redakteur te gee indien die voorlegging weens enige rede onttrek word.

7. Tensy reelings vooraf met die redakteur getref is, mag 'n artikel (insluitend voetnotas en opsomming) nie 8000 woorde oorskry nie. Ender bydraes mag nie 5500 woorde oorskry nie.

8. Tegniese riglyne vir outeurs is op die webblad van die Fakulteit Regsgeleerdheid van die Universiteit van Pretoria bekikbaar: http://www.up.ac.za/academic/law/dejure.

Editorial Policy

1. *De Jure* has the promotion of legal science and a critical and analytical approach to law as its main objectives and towards this aim publishes original contributions of a high academic standard.

2. *De Jure* is a national and general legal journal to which academics, members of the judiciary and members of the different legal professions may contribute. No preference is given to authors from any particular institution. The decision whether to publish any submission depends on whether it meets the high quality standards of *De Jure* and whether space is available for publication.

3. Contributions in both English and Afrikaans are published and may consist of articles, notes, discussions of recent cases and book reviews. A translated title and a brief summary of approximately 300 words in Afrikaans must accompany an article written in English. In the case of Afrikaans articles a similar requirement applies regarding an English title and summary.

4. In order to be considered for publication a contribution must be the result of original research by the author(s), meet with all applicable legal principles in respect of publication (such as copyright, etcetera), contribute something sufficiently new to the existing legal literature and conform to the linguistic, technical and stylistic requirements for publications in *De Jure*. Authors are personally responsible to ensure that their submissions meet all these requirements.

5. Submissions are accepted for consideration only on the basis that while the editorial committee makes the final decision on publication, submissions will be subjected to appropriate peer and expert review, as well as review by members of the advisory committee when necessary. The editorial committee further reserves the right to edit all submissions accepted for publication in terms of the editorial policy, as well as to shorten submissions if necessary.

6. Manuscripts may only be submitted in electronic format (utilising MS Word) through an e-mail to the editor or on an electronic disc delivered to the editors office. In addition, authors

(a) must supply their relevant contact particulars, especially e-mail address(es) and telephone numbers;
(b) guarantee that they are legally entitled to have the full submission published and that it, or a part of it, has not been published elsewhere before;
(c) disclose whether it, or a part of it, has been submitted to any other journal for publication; and
(d) undertake to give reasonable notice to the editor if the submission is withdrawn for any reason.

7. Unless prior arrangements have been made with the editor, an article (including footnotes and the summary) may not exceed 8000 words and other contributions may not exceed 5500 words.

8. Technical guidelines to authors are available on the website of the Faculty of Law of the University of Pretoria: http://www.up.ac.za/academic/law/dejure